Professionals Center for Business Research

Research Article Volume 2 - February 2015 (02)

Commercial Chatbot: Performance Evaluation, Usability Metrics and Quality Standards of Embodied Conversational Agents

Karolina Kuligowska[1]

[1]Department of Information Systems and Economic Analysis, Faculty of Economic Sciences, University of Warsaw, Warsaw, Poland

Abstract: The aim of this paper is to explore commercial applications of chatbots, as well as to propose several measurement metrics to evaluate performance, usability and overall quality of an embodied conversational agent. On the basis of these metrics we examine existing Polish-speaking commercial chatbots that a) work in the B2C sector, b) reach the widest possible range of users, and c) are presumably the most advanced commercial deployments of their creators. We analyse various aspects of functioning of each embodied conversational agent: visual look, form of implementation on the website, speech synthesis unit, built-in knowledge base (with general and specialized information), presentation of knowledge and additional functionalities, conversational abilities and context sensitiveness, personality traits, personalization options, emergency responses in unexpected situations, possibility of rating chatbot and the website by the user. Our study reveals the current condition of Polish market of commercial virtual assistants and emphasizes the importance of a multidimensional evaluation of any commercial chatbot deployment.

Keywords: commercial chatbot, virtual assistant, virtual agent, embodied conversational agent, chatbot developer

1 Introduction

In 2003, the first Polish commercial chatbot was launched. He had an appearance of a smiling man, but it was only a static photography. He was able to converse with users, however only in a silent, textual way. Eventually, his knowledge was limited, imprecise and quite vague.

For over 10 years the market of humanlike online virtual assistants has been growing fast. More and more sophisticated technological solutions were implemented, as well as innovative designs of knowledge bases were created, and many new functionalities were added. Chatbot developers, i.e. companies that produce virtual assistants, seemed to spring up like mushrooms. Some of these companies have already fallen, other refocused on another business field or merged with bigger foreign companies. Nevertheless, a few of them flourished with success in this futuristic field of conversational artificial intelligence. What is the current condition of Polish market of commercial virtual assistants? Do virtual agents meet customers' expectations of quick, pleasant and professional services?

The aim of this paper is to explore commercial applications of chatbots by applying several measurement metrics to evaluate performance, usability and overall quality of Polish-speaking commercial chatbots. The paper is organized as follows. Section 2 presents substantiation of list of chatbots, that were subject of examination in our analysis. Section 3 introduces several measurement metrics and reviews various aspects of functioning of each virtual assistant. Section 4 discusses obtained results and provides a summary of our findings. Finally, the conclusions due to this paper are considered in Section 5.

2 Existing Polish-Speaking Commercial Chatbots

All currently working commercial chatbots, that are able to converse in Polish, were created by six following chatbot developers (listed in alphabetical order): Artificial Solutions[1], Denise Systems[2], Fido intelligence[3], InteliWISE[4], Onlinetools.p[5], Stanusch

[1] http://www.artificial-solutions.com (July 2014)
[2] http://denisesystems.pl (July 2014)
[3] http://fidointelligence.pl (July 2014)

Published at: http://pcbr.net/issue/0202/
Article Number: 0202.22; Online ISSN: 2409-9783

Karolina Kuligowska (Correspondence)
kkuligowska@wne.uw.edu.pl

Technologies[6]. Virtual assistants produced by these companies are deployed in a variety of industries and business fields. Unfortunately, it is not possible to find one particular industry, within which all the above-mentioned companies would be represented by their commercial chatbot deployments. Since comparing all solutions within one industry field is not doable, therefore we chose and depicted from every company one representative commercial deployment of a virtual assistant.

In the first step, we identified all active Polish-speaking commercial chatbots publicly available on the Internet. The fixed period, for which a chatbot is functioning on a website, is often subject to an expiration. Not surprisingly then, some companies have at the present time just one virtual assistant that can be considered as a representative implementation (Artificial Solutions, Fido intelligence). Other companies (Denise Systems, InteliWISE, Onlinetools.pl, Stanusch Technologies) have more than one currently working representative implementation (to be exact: 3, 5, 2, and 17 respectively). All identified Polish-speaking commercial chatbots are shown in Table 1

4 http://www.inteliwise.com/pl (July 2014)
5 http://www.onlinetools.pl/ (July 2014)
6 http://www.stanusch.com/pl/ (July 2014)

Table 1. List of existing Polish-speaking commercial chatbots (July 2014) - sorted by company name in alphabetical order

Chatbot developer	Chatbot name	Works on behalf of	Website
Artificial Solutions	Ania	IKEA	http://www.ikea.com/pl/pl
Denise Systems	Zosia	Villa Pan Tadeusz	http://www.villapantadeusz.pl
Denise Systems	Julia	Hotel Chrobry	http://www.hotelchrobry.pl
Denise Systems	Kasia	SmartProgress	http://www.smartprogress.pl
Fido intelligence	Karen	WSHiFM	http://www.wshifm.edu.pl
InteliWISE	Wirtualny Doradca	PayU	http://www.payu.pl/pomoc
InteliWISE	Wirtualna Doradczyni	NaviExpert	http://www.naviexpert.pl/pomoc/wirtualny-doradca.html
InteliWISE	Mubu	Mobile Interactive	http://www.mubu.pl
InteliWISE	Wirtualny Doradca	Warta	http://www.warta.pl
InteliWISE	Beata	Centrum Pomocy CEIDG	https://prod.ceidg.gov.pl
Onlinetools.pl	Ewa/Adam	Kredytum.pl	http://kredytum.pl
Onlinetools.pl	Szkoleniowiec BHP	Kendos	http://bsdf.pl
Stanusch Technologies	Ewa	Orange	http://www.orange.pl/obsluga_klienta_indywidualnego.phtml
Stanusch Technologies	Iza	BNP Paribas	http://www.bnpparibas.pl
Stanusch Technologies	Anna	InPost	http://www.inpost.pl
Stanusch Technologies	Magda	Gaspol	http://www.dobrzeogrzanydom.pl
Stanusch Technologies	Magda	Matras	http://www.matras.pl
Stanusch Technologies	Magda	Virbac	http://www.virbac.pl
Stanusch Technologies	Promozonka	Promozone	http://www.promozone.pl
Stanusch Technologies	Milena	AutoCentrum.pl	http://www.autocentrum.pl
Stanusch Technologies	Paula	Electronic-Marketing	http://www.e-marketing.pl
Stanusch Technologies	Anna/Marek	Best Consulting Plus	http://www.opti.pl
Stanusch Technologies	Monika /Krzysztof	PUE ZUS	http://pue.zus.pl
Stanusch Technologies	Ewa	PUP Piotrków Trybunalski	http://www.ibe.puppiotrkow.pl
Stanusch Technologies	Ewa	Europejskie Stowarzyszenie Czyste Leczenie	http://czysteleczenie.pl

Stanusch Technologies	Klaudia	AGH Kraków	http://www.dsp.agh.edu.pl
Stanusch Technologies	Wincenty	UE Katowice	http://inzynieriawiedzy.pl
Stanusch Technologies	Ewa	Progra	http://www.awatar.progra.pl
Stanusch Technologies	Ewa	Projekt Koordynacja 3.0	http://www.koordynacja.progra.pl

Source: own elaboration.

In the second step, we limited the list of representative solutions to virtual assistants working in the B2C sector. Thus we rejected chatbots from virtual offices of nonprofit organizations and virtual experts representing government organizations. Motivation for this was the assumption that private entities (companies as well as private universities) care most on the highest rate of return on investment in deploying a virtual assistant, while maximizing profits and operational efficiencies, and minimizing related costs.

In the third step, we selected chatbot deployments that reach the widest possible range of users - potential customers of the company. Thus we rejected chatbots working on behalf of small local companies, interacting with a limited number of users.

In the end, from the remaining set we picked out implementations considered as being the best or most advanced representative commercial deployments of a given company. Hence, the ultimate list of chatbots that were subject of examination in our study, is presented in Table 2.

Table 2. List of examined virtual assistants - sorted by company name in alphabetical order

Chatbot developer	Chatbot name	Works on behalf of	Website
Artificial Solutions	Ania	IKEA	http://www.ikea.com/pl/pl/
Denise Systems	Zosia	Villa Pan Tadeusz	http://www.villapantadeusz.pl/
Fido intelligence	Karen	WSHiFM	http://www.wshifm.edu.pl/
InteliWISE	Wirtualny Doradca	PayU	http://www.payu.pl/pomoc
Onlinetools.pl	Ewa/Adam	Kredytum.pl	http://kredytum.pl
Stanusch Technologies	Ewa	Orange	http://www.orange.pl/obsluga_klienta_indywidualnego.phtml

Source: own elaboration.

3 Quality Components and Their Evaluation

Ten various aspects of functioning of each virtual assistant were subjected to our comprehensive analysis. We examined the visual look of the chatbot, its form of implementation on the website, and the speech synthesis unit. In order to evaluate the built-in knowledge base (containing general and specialized information) in a comprehensive, reliable and accurate way, each virtual assistant was asked the same set of questions. By that means we also studied available forms of presentation of knowledge and additional functionalities performed by each chatbot. During interviews we examined conversational abilities, language skills and context sensitiveness of a chatbot linguistic module. Furthermore, we evaluated personality traits, personalization options, and emergency responses in unexpected situations. Finally, we assessed the possibility of rating chatbot and the website by the user.

Proposed evaluation includes quality components of every virtual assistant. It is adapted to the commercial chatbot deployment and it reflects user needs, preferences and expectations in particular [Abu Shawar, Atwell, 2007:96] [Knoppel et al., 2008:113]. Moreover, from a commercial applications evaluation standpoint, a focus must be put not only on the end-user satisfaction, but also on the benefit of the website owner on behalf of which the virtual assistant is working [Friedlein 2003:269-340] [Gaudiano, Kater,

2000:124].

In order to evaluate quality components, we used standard measurement tool of 1-5 rating scale, denoting: 1-very poor, 2-poor, 3-satisfactory, 4-good, and 5-very good rating. Assessment was granted each time on the basis of the information resulting from previous research results, literature review and best business practices. Some components were divided into five evaluated parts, where rating was given on a basis of a sum of earned points. Very simple and transparent measures were used, therefore there was no need to transform data for advanced statistical analysis. In the final section 4, we calculated a simple average of all evaluated points, which gives an insight into overall quality of each commercial chatbot implementation.

3.1 Visual Look of the Chatbot

The book is often judged by its cover. Therefore an outward appearance of a virtual assistant constitutes an important factor that influences the quality standard of its implementation. Findings on chatbot visualization emphasize the importance of its visual look [Haake 2009:35-56], and clearly point to the benefit to using faces (human and animated) [Berry et al., 2004:34]. A chatbot visualization that resembles a living person, increases user involvement and willingness to start a conversation [Van Vugt et al., 2010:21-22]. It is as well recommended to achieve the consistency between verbal and non-verbal behaviours of conversational animated agents [Berry et al., 2004:10]. In a word, chatbots should demonstrate ability of rich social interactions through natural conversation, flexible expressiveness, and convergent gestural behavior [Kopp et al., 2009:508].

It would seem obvious that, with the available advanced technology, video sequences constitute a standard realization of a commercial chatbot visualization. However, it turns out that even today exist bodiless virtual assistants (PayU) or those represented by a modest cartoon-like animation (IKEA). Yet most visualizations are high quality video sequences depicting a living person that is speaking, gesturing and behaving accordingly to displayed utterances (Villa Pan Tadeusz, WSHiFM, Kredytum.pl, Orange).

Video sequences depicting a living person as chatbot visualization were assessed as very good. Cartoon-like animation constitutes still good appearance, while disembodied chatbot was rated as very poor, as presented in Table 3.

Table 3. Chatbot visualization and its evaluation on a scale of 1 (very poor) to 5 (very good)

Chatbot name (works on behalf of)	Visual look of the chatbot	Rating
Ania (IKEA)	cartoon-like animation	4
Zosia (Villa Pan Tadeusz)	video sequences depicting a living person	5
Karen (WSHiFM)	video sequences depicting a living person	5
Wirtualny Doradca (PayU)	disembodied	1
Ewa/Adam (Kredytum.pl)	video sequences depicting a living person	5
Ewa (Orange)	video sequences depicting a living person	5

Source: own elaboration.

3.2 Form of Implementation on the Website

One of the key aspects of embedding a virtual assistant on the website is the visibility [Gaudiano, Kater, 2000:121] [Hsu 2011:60-71]. A company that purchases a chatbot is rarely willing to completely rebuild its existing web service, therefore forms of implementation of virtual assistants tend to be quite varied.

Virtual assistants are embedded in the form of a floating window (IKEA), a built-in window (WSHiFM, PayU), and a pull-out side tab (Villa Pan Tadeusz, Kredytum.pl). In some cases, there exists a flexible combination of two solutions: a built-in window on main website and a pull-out side tab appearing on the subpages (Orange).

Flexible combination of a built-in window and a pull-out side tab as a form of chatbot implementation was assessed as very good. Pull-out side tab constitutes still good solution, while fixed built-in window or small floating window were rated only as satisfactory, as presented in Table 4.

Table 4. Form of implementation of chatbot on the website and its evaluation on a scale of 1 (very poor) to 5 (very good)

Chatbot name (works on behalf of)	Form of implementation on the website	Rating
Ania (IKEA)	floating window	3
Zosia (Villa Pan Tadeusz)	pull-out side tab	4
Karen (WSHiFM)	built-in window	3
Wirtualny Doradca (PayU)	built-in window	3
Ewa/Adam (Kredytum.pl)	pull-out side tab	4
Ewa (Orange)	flexible combination	5

Source: own elaboration.

3.3 Speech Synthesis Unit

One of the quality components of a virtual assistant is the Text-To-Speech module that converts written text into a synthetic speech [Van Deemter et al., 2008:1239-1243]. Research shows that embodied conversational agents can use their synthesized voice in order to increase users trust [Elkins, Derrick, 2013:910-912].

Some virtual agents realize their conversations in silence (IKEA, PayU). In contrast, others speak with a human voice, using a speech synthesis unit (Villa Pan Tadeusz, WSHiFM, Kredytum.pl, Orange). Speech synthesis module may be standard, delivered by an external provider (Villa Pan Tadeusz, WSHiFM, Kredytum.pl) or dedicated to a particular implementation, specially prepared unique voice (Orange).

Speech synthesis unit with a unique custom voice equipped with shutdown option was assessed as very good. Standard Text-to-Speech solution equipped with shutdown option constitutes still good speech module, while standard voice without shutdown option was evaluated only as satisfactory. Absence of voice was rated as very poor, as presented in Table 5.

Table 5. Speech synthesis unit and its evaluation on a scale of 1 (very poor) to 5 (very good)

Chatbot name (works on behalf of)	Speech synthesis unit	Rating
Ania (IKEA)	voiceless	1
Zosia (Villa Pan Tadeusz)	standard voice, no shutdown option	3
Karen (WSHiFM)	standard voice, shutdown option	4
Wirtualny Doradca (PayU)	voiceless	1
Ewa/Adam (Kredytum.pl)	standard voice, shutdown option	4
Ewa (Orange)	unique voice, shutdown option	5

Source: own elaboration.

3.4 Knowledge Base

The knowledge base stores for a virtual assistant all data, information and knowledge about reality, and hence it has a fundamental meaning for the functioning of the virtual assistant [Gaudiano, Kater, 2000:122-123]. It constitutes the essence of chatbot's existence and is the main reason for its creation. A conversational agent should be able to talk about any topic. Hence flows the challenge of evaluating knowledge base: sufficiently comprehensive knowledge base includes both on-domain and off-domain answers [Kenny et al., 2007:202-203] [Sjödén et al., 2011:123-124]. Therefore knowledge base evaluation is divided into two parts: general information topics and specialized business knowledge

- **Knowledge Base: Basic Knowledge**

Examined virtual assistants were asked several questions involving knowledge at the level of an elementary school graduate. Sample questions: who are you, how old are you, what's your name, who is your father, how are you, today's date, how much is 3 times 3, who is the president of Poland, display information on capital cities, solar system and planets, etc.

All Polish-speaking commercial chatbots have general information about themselves and the surrounding world, on the basic level. However, only some of them know who created them (IKEA, Villa Pan Tadeusz, WSHiFM, Kredytum.pl, Orange), and even less are able to determine the start date of their functioning on the website (Villa Pan Tadeusz, WSHiFM, Orange). Few virtual assistants provide information about the current date and time (WSHiFM, Kredytum.pl, Orange). Whereas knowledge concerning geography, history, astronomy, the meaning of Polish names and the ability to do simple mathematical cal-

culations has only one commercial chatbot (Orange).

In order to make the results comparable, we presumed that each chatbot should be able to answer 5 simple key questions that check the built-in basic knowledge. These questions are as follows:
Q1 – what is your name, Q2 – what can you do, Q3 – who do you work for, Q4 – what time is it, Q5 – what is the capital of Poland.

For each correct answer the virtual assistant obtained one point. In case of absence of the correct answer, if the context and meaning of the question were correctly understood, the virtual assistant received a half-point. Lack of any answer reasonably consistent with the dialogue scope resulted in the lack of points. Correct answers for basic knowledge topics (general information), sum of earned points and their rating are presented in Table 6.

Table 6. Basic knowledge questions and answers, and their evaluation on a scale of 1 (very poor) to 5 (very good)

Chatbot name (works on behalf of)	**Q1**	**Q2**	**Q3**	**Q4**	**Q5**	**Sum**	**Rating**
Ania (IKEA)	1	1	1	0.5	0.5	4	4
Zosia (Villa Pan Tadeusz)	1	1	0.5	0.5	0	3	3
Karen (WSHiFM)	1	1	0.5	1	0.5	4	4
Wirtualny Doradca (PayU)	1	1	0	0	0	2	2
Ewa/Adam (Kredytum.pl)	1	1	0.5	1	0.5	4	4
Ewa (Orange)	1	1	1	1	1	5	5

Source: own elaboration.

- **Knowledge Base: Specialized Knowledge**

All Polish-speaking commercial chatbots have specialized knowledge about the represented company, its products and services. Yet, the effective use of this knowledge is hardly comparable, because it depends heavily on the industry branch within which the given company is represented by its virtual assistant.

Despite the above-mentioned difficulties of comparative studies, we presumed that each commercial chatbot should be able to answer 5 key questions that check the built-in specialized knowledge. The questions were composed as objectively and broadly as it was possible, without favoring any of the industries. These questions are as follows: Q1 – what is the offer/display product catalog, Q2 – how much does it cost/display prices, Q3 – how do I get a discount/display rebates, Q4 – how do I contact the office/display contact information, Q5 – does your company have any achievements/display success stories. Questions Q1, Q2, Q3 concern specialized knowledge about products and services offered. Question Q4 inspects necessary knowledge of contact data of the represented company. Question Q5 verifies the advanced knowledge about the company, namely about its achievements.

For each answer containing exhaustive information regarding the question, the virtual assistant obtained one point. In case of providing a partial answer to the question, if the context and meaning of the question were correctly understood, the virtual assistant received a half-point. Lack of any answer reasonably consistent with the dialogue scope resulted in the lack of points. Correct answers for specialized knowledge topics (competence and expertise), sum of earned points and their rating are presented in Table 7.

Table 7. Specialized knowledge questions and answers, and their evaluation on a scale of 1 (very poor) to 5 (very good)

Chatbot name (works on behalf of)	Q1	Q2	Q3	Q4	Q5	Sum	Rating
Ania (IKEA)	1	1	0	1	0.5	3.5	4
Zosia (Villa Pan Tadeusz)	0.5	1	0.5	1	1	4	4
Karen (WSHiFM)	1	1	1	0.5	1	4.5	5
Wirtualny Doradca (PayU)	0.5	1	1	1	0.5	4	4
Ewa/Adam (Kredytum.pl)	1	0	0	1	0.5	2.5	3
Ewa (Orange)	1	1	1	1	1	5	5

Source: own elaboration.

3.5 Presentation of Knowledge and Additional Functionalities Performed by Chatbots

In most embodied conversational agents the only autonomous functionality used is that of initiation of a conversation. Yet, capability to trigger behavior and use additional functionalities in a context sensitive way constitutes one of the significant characteristics of commercial chatbots [Cassell et al., 1999b:30] [Gerhard 2006:3] [Kumar, Rosé, 2009:5-7]. Therefore evaluation of the exceptional form of presentation of knowledge, as well as additional functionalities performed by every virtual agent, and other available ways of facilitating user navigation on the website, completes the competence assessment.

All tested virtual assistants display, within their utterances, embedded clickable links. The user can click them and get answers immediately, without having to write questions manually. Most chatbots display such embedded links dynamically, as tagged words of phrases available to be clicked by user in order to continue the conversation (IKEA, Villa Pan Tadeusz, WSHiFM, PayU, Orange). Others use a static list of links displayed below the chatbot utterance (Kredytum.pl).

However, some users are not likely to click on embedded links. In such cases a possibility of autonomous acting is crucial for a commercial chatbot. Autonomous ability of dynamically loading new subpages or tabs in background or in a new window during a conversation exhibit half of the analyzed virtual assistants (IKEA, Villa Pan Tadeusz, WSHiFM).

Interactive connection to an external database concerns the product catalog in one case (IKEA), and in the second case - intranet and the possibility of trying to find a pertinent answer by executing web mining, i.e. searching information on the web (Orange).

The button "Back" in the form of a triangle or an arrow (Kredytum.pl, Orange), as well as scrolling the chat history (WSHiFM, PayU, Kredytum.pl) allows users to view previously displayed answers.

Some chatbots are able to explain their functioning after entering the term "Help" or when pressing the corresponding button "Info", "i" or "?" (IKEA, Kredytum.pl).

Button "Home" accelerates the return to the main menu, where the most important information for the customers are listed (Kredytum.pl, Orange). Possible forms of presentation of knowledge and additional functionalities performed by virtual assistants are presented in Table 8.

Table 8. Presentation of knowledge and additional functionalities performed by chatbots, and their evaluation on a scale of 1 (very poor) to 5 (very good)

Chatbot name (works on behalf of)	Autonomous dynamic loading new subpages or tabs in background or in a new window	Interactive connection to an external database	"Back" button or scrolling the chat history	Term "Help" or "Info", "i", "?" button	"Home" button - return to the main menu	Sum	Rating
Ania (IKEA)	+	+	-	+	-	3	3
Zosia (Villa Pan Tadeusz)	+	-	-	-	-	1	1
Karen (WSHiFM)	+	-	+	-	-	2	2
Wirtualny Doradca (PayU)	-	-	+	-	-	1	1
Ewa/Adam (Kredytum.pl)	-	-	+	+	+	3	3
Ewa (Orange)	-	+	+	-	+	3	3

Source: own elaboration.

3.6 Conversational Abilities, Language Skills and Context Sensitiveness

Chatbot's job is to generate clear and coherent expressions, while taking into account proper social communication behavior. The biggest challenge in designing a chatbot architecture is inventing the mechanism of dialogue context detection, which makes the chatbot keeping pace with a constantly changing topic of conversation. Conversational abilities involve not only natural language processing, but also speaking a large set of utterances, and managing discourse by combining text from different categories of topic groups to produce the final response [Kenny, Parsons, 2011:262-263] [Stavropoulou et al., 2011:336]. Language skills and context sensitiveness require the integration of the conversational behaviors such as giving feedback, taking turn, and repairing of dialogue [Cassell et al., 1999a:524-525] [Heudin 2010:5] [Traum 2008:48-52].

Yet conversational skills are not a strong point of every commercial chatbot. Some virtual assistants understand small amount of utterances and they try to overcome this issue by invariably displaying long lists of suggested useful links (PayU, Kredytum.pl). Others suggest a conversation topic, for example "do you know what is the offer of a student careers office", and when the user asks subsequently "so what is the offer", they lose context and they suddenly retreat by replying "I'm sorry but at the moment I don't know what's the offer" (WSHiFM). Another inexplicable conversational reaction is the answer "I am not afraid of anything" as a response to the user's previous statement "I am glad" (Kredytum.pl).

Moreover, some chatbots present embedded clickable links, that mislead a user. Clicking the embedded link "what is included in the price" leads to a strange answer "I can't even guess" (Villa Pan Tadeusz), while following the link "collaboration with foreign universities" and "academic IT center" leads to blank subpages (WSHiFM). Oddly enough, clicking the link "contact" loads a subpage with latest news instead of a subpage containing contact data (WSHiFM). Finally, functioning of some virtual agents resembles more a search engine of frequently asked questions than a conversational agent (PayU).

However, there are also chatbots that lead a coherent dialogue (IKEA, Villa Pan Tadeusz, Orange), handle complex user input and take control of the dialogue introducing topics, requesting clarifications, and sug-

gesting to search for a pertinent answer within the website and intranet in case of lack of information in a specific domain (Orange).

Skills of leading a coherent dialogue, understanding complex expressions, and capabilities of repairing the dialogue were assessed as very good. Leading a coherent dialogue constitutes still a good conversational ability, while understanding small amount of utterances was rated only as satisfactory, as presented in Table 9.

Table 9. Conversational abilities, language skills and context sensitiveness, and their evaluation on a scale of 1 (very poor) to 5 (very good)

Chatbot name (works on behalf of)	Conversational abilities, language skills and context sensitiveness	Rating
Ania (IKEA)	leads a coherent dialogue	4
Zosia (Villa Pan Tadeusz)	leads a coherent dialogue	4
Karen (WSHiFM)	understands small amount of utterances	3
Wirtualny Doradca (PayU)	understands small amount of utterances	3
Ewa/Adam (Kredytum.pl)	understands small amount of utterances	3
Ewa (Orange)	leads a coherent dialogue, understands complex expressions, repairs the dialogue	5

Source: own elaboration.

3.7 Personality Traits

In order to become believable in the eyes of users, commercial chatbots not only must be equipped with expertise, but also with the expression of personality. It is therefore important to incorporate a number of psychological layers to the knowledge base of a virtual assistant, including personality traits, biographical facts and expressed emotions [Callejas et al., 2011b:213-215] [Heudin 2011:252-256] [Kasap, Magnenat-Thalmann, 2007:9-12] [Knoppel et al., 2008:115] [Loyall, Bates, 1997:111] [Ortony 2003:201-205].

Some Polish-speaking commercial chatbots have a rudimentary outline of personality (PayU). On the contrary, there are several virtual assistants containing a very rich personality expressed throughout information about their self-being, such as age, gender, interests etc., as well as psychological outline, specifying personality that is coherent with emotional reactions (IKEA, Villa Pan Tadeusz, WSHiFM, Kredytum.pl, Orange).

Very rich personality was assessed as very good, while rudimentary outline of personality was rated only as satisfactory, as presented in Table 10.

Table 10. Personality traits and their evaluation on a scale of 1 (very poor) to 5 (very good)

Chatbot name (works on behalf of)	Personality traits	Rating
Ania (IKEA)	very rich personality	5
Zosia (Villa Pan Tadeusz)	very rich personality	5
Karen (WSHiFM)	very rich personality	5
Wirtualny Doradca (PayU)	rudimentary outline of personality	3
Ewa/Adam (Kredytum.pl)	very rich personality	5
Ewa (Orange)	very rich personality	5

Source: own elaboration.

3.8 Personalization Options

Delivering a consistent personalized experience is one of the largest CRM challenges [Friedlein 2003:241-252]. Personalization options have a significant positive effect on users' assessment of the quality of interaction with a commercial chatbot. Users build up impressions of a virtual agent by confirming whether its various characteristics match their preferences [Thomas, Fischer, 1997:58]. Even if it is just an illusion of customization, it can still outweigh the perceived quality of the given virtual assistant. Research in this matter shows, that when allowed to choose the characteristics and appearance of an embodied conversational agent, users viewed it as more likable, more trustworthy and more useful [Xiao et al., 2007:1299-1302].

One chatbot provides the users with a personalization option allowing the change of the gender of its visualization (Kredytum.pl). One virtual assistant is able to remember and recall the name of the user (WSHiFM), skilfully adjusting the displayed text

according to the gender of its interlocutor. Viewing the conversation history during the talk is possible in some cases (WSHiFM, PayU, Kredytum.pl, Orange), while printing the whole dialogue text or sending the conversation history via an email is offered by only one commercial chatbot (PayU). Also only one chatbot can recognize the subpage that the user is browsing, which results in displaying the content adequate to the given subpage, making it easier to navigate and to find the needed information on a website (Orange).

Table 11. Presentation of personalization options performed by virtual assistants, and their evaluation on a scale of 1 (very poor) to 5 (very good)

Chatbot name (works on behalf of)	Change of the gender of chatbot visualization	Recalling the name of the user	Viewing the conversation history during the talk	Access to the conversation history (print or send via email)	Recognize the subpage browsed	Sum	Rating
Ania (IKEA)	-	-	-	-	-	0	1
Zosia (Villa Pan Tadeusz)	-	-	-	-	-	0	1
Karen (WSHiFM)	-	+	+	-	-	2	2
Wirtualny Doradca (PayU)	-	-	+	+	-	2	2
Ewa/Adam (Kredytum.pl)	+	-	+	-	-	2	2
Ewa (Orange)	-	-	+	-	+	2	2

Source: own elaboration.

3.9 Emergency Responses in Unexpected Situations

Knowledge, behavior and stress-resistance of virtual assistants are often rudely probed by users. They mislead chatbots, ask about abstract concepts or express derogatory attitudes. Many studies reveal that aggression, verbal abuse and sexual harassment directed towards a chatbot tend to occur frequently during conversations [Brahnam 2006:13-16] [De Angeli 2006:21-24] [De Angeli, Carpenter, 2005:21-25]. Commercial virtual assistant should be able to intelligently, diplomatically and patiently respond to such emergency situations. Chatbot should also recognize any typos, misspellings or colloquialisms used in the conversation.

In case when the user's statement is difficult to understand or hard to interpret, virtual assistant has to handle somehow this unexpected situation. There are virtual assistants that understand large amount of utterances and when they realize lack of information on a specific question, they try many inventive ways to overcome their ignorance. For example, in case of hesitation, they make sure if their answer was helpful (IKEA) or they offer to search information within the website and intranet (Orange) demonstrating web mining capabilities [Heudin 2010:7-10] [Millet, Heudin, 2007:3-4].

Alternative diplomatic riposte, when the virtual assistant realizes lack of knowledge, is to take the initiative and gather customer data, for instance by asking users questions concerning their age, education, occupation, monthly income, size of the place of residence, and name of the city of origin (Villa Pan Tadeusz). Another clever solution is to behave proactively by changing a topic of conversation and displaying important information every once in a while (WSHiFM, Orange).

Typos, misspellings and various errors commonly appear in users' statements. A big challenge for each chatbot is to overcome these mistakes. Of course it is not possible to recognize all typos. Though, at least in cases of simple words and phrases, chatbot should be able to understand user's utterance without asking to reformulate the entered sentence. Absolutely the best in dealing with typos, misspellings and various er-

rors, are only two chatbots (IKEA, Orange). Other three are able to recognize distorted statements, yet it happens very rarely (Villa Pan Tadeusz, WSHiFM, PayU). One virtual assistant does not detect typos nor misspellings at all (Kredytum.pl).

In case of recognizing offensive statements and insults, the most appropriate behavior for a commercial chatbot is never to get provoked nor to rise to the level of anger or frustration of the customer. Defusing hostility by controlling emotional reactions, modeling professional behavior and refocusing on problem solving is recommended by a number of CRM methods of handling verbally abusive customers [Brahnam 2005:64]. All what a virtual assistant can do is to ignore calumnies or take pity on the user and softly return to the main topic of conversation (IKEA, Villa Pan Tadeusz, PayU, Kredytum.pl, Orange). Inefficient way of coping with insults is to fuel the fire with returning the same behavior. Commercial chatbot that shows resentment to the extent of finishing the conversation (WSHiFM) and depriving the user of the possibility of asking further questions (WSHiFM), treats the user disrespectfully and risks tarnishing the image of the company [Brahnam 2005:66-67]. Inept responses to customer abuse make the website a hostile environment, discourage users from future visits [Berry et al., 2004:37] [Stern 2003:336] [Van Vugt et al., 2010:19-20] and can sour the customer relationship.

It happens that a chatbot recognizes the question about speaking a foreign language such as "do you speak English" (Villa Pan Tadeusz, WSHiFM, Orange) or similarly in French, Spanish, and German (IKEA, WSHiFM). Some of them inform in Polish, that they do not speak foreign languages (IKEA, Villa Pan Tadeusz). Other virtual assistants display simple answers in English (WSHiFM, Orange) or similarly in French, Spanish, and German (WSHiFM). However, only one virtual agent presents a unique ability of translating simple words from English into Polish (Orange) - users just need to write a command "translate" (in Polish) followed by an English term, and then click enter[7].

[7] For example, in order to translate the term "virtual agent", enter: *tłumacz virtual agent*.

Table 12. Presentation of handling emergency responses in unexpected situations, and their evaluation on a scale of 1 (very poor) to 5 (very good)

Chatbot name (works on behalf of)	Overcoming ignorance	Overcoming typos and misspellings	Overcoming insults and humiliation	Foreign languages recognition	Translating English words into Polish	Sum	Rating
Ania (IKEA)	+	+	+	+	-	4	4
Zosia (Villa Pan Tadeusz)	+	+	+	+	-	4	4
Karen (WSHiFM)	+	+	-	+	-	3	3
Wirtualny Doradca (PayU)	-	+	+	-	-	2	2
Ewa/Adam (Kredytum.pl)	-	-	+	-	-	1	1
Ewa (Orange)	+	+	+	+	+	5	5

Source: own elaboration.

3.10 Possibility of Rating Chatbot and the Website by the User

Conversation with a virtual assistant serves not only as the presentation of the business offer to a potential customer. The added value for the chatbot owner is the user feedback on the chatbot and/or on the website [Callejas et al., 2011a:14] [Knoppel et al., 2008:117] [Kowalski et al., 2013:271], as well as customers' willingness to recommend a product or service to someone else.

Some virtual assistants make sure during the conversation whether their answer was helpful for the user and they ask for a comment (IKEA) or they ask for an assessment of various features (Villa Pan Tadeusz). Others provide users with separate five-star quality rating system, either for evaluating single answers (PayU), or for measuring user overall satisfaction (Orange). In turn, other chatbots do not ask for any feedback (WSHiFM, Kredytum.pl).

Possibility of rating overall satisfaction (whether with five-star rating system or during dialogue) was assessed as very good. Possibility of evaluating single answers (whether with five-star rating system or during dialogue) was also assessed as very good. No interest in user feedback was rated as very poor, as presented in Table 13.

Table 13. Possibility of rating chatbot and the website by the user and its evaluation on a scale of 1 (very poor) to 5 (very good)

Chatbot name (works on behalf of)	Possibility of rating chatbot and the website by the user	Rating
Ania (IKEA)	single answer, ask for assessment	5
Zosia (Villa Pan Tadeusz)	overall satisfaction, ask for assessment	5
Karen (WSHiFM)	no interest in user feedback	1
Wirtualny Doradca (PayU)	single answer, five-star rating system	5
Ewa/Adam (Kredytum.pl)	no interest in user feedback	1
Ewa (Orange)	overall satisfaction, five-star rating system	5

Source: own elaboration

4 Discussion

We conducted our study to explore commercial applications of chatbots. We defined ten pertinent components that determine the quality of every commercial chatbot deployment: visual look, form of implementation on the website, speech synthesis unit, built-in knowledge base (with general and specialized information), presentation of knowledge and additional functionalities, conversational abilities and context sensitiveness, personality traits, personalization options, emergency responses in unexpected situations, possibility of rating chatbot and the web-

site by the user.

Each quality component was evaluated on a scale of 1 (very poor) to 5 (very good). Some components were divided into five evaluated parts, where rating was given on a basis of a sum of earned points. Table 14 provides a detailed overview of the ratings of every component and functionality. For each commercial chatbot implementation we also present an average of all earned points, rounded rating and resulting overall quality.

Ten-year period of market development of commercial chatbots may be assessed positively, since the implementations rated as very good (Orange), good (IKEA, Villa Pan Tadeusz), and satisfactory (WSHiFM, PayU, Kredytum.pl) outweigh. Virtual assistants that gained the highest rating, and may be considered as three market's winners, are as follows: Ewa working on behalf of Orange (very good overall quality), Ania working on behalf of IKEA (good overall quality), and Zosia working on behalf of Villa Pan Tadeusz (good overall quality). The producers of these commercial chatbots are respectively: Stanusch Technologies, Artificial Solutions, and Denise Systems.

Table 14. Overview of the ratings of each functionality of the commercial chatbot on a scale of 1 (very poor) to 5 (very good)

Chatbot name (works on behalf of)	Visual look	Form of implementation on the website	Speech synthesis unit	Knowledge base (basic knowledge)	Knowledge base (specialized knowledge)	Presentation of knowledge and additional functionalities	Conversational abilities, language skills and context sensitiveness	Personality traits	Personalization options	Emergency responses in unexpected situations	Possibility of rating chatbot and the website by the user	AVERAGE	RATING	OVERALL QUALITY
Ania (IKEA)	4	3	1	4	4	3	4	5	1	4	5	3,5	4	good
Zosia (Villa Pan Tadeusz)	5	4	3	3	4	1	4	5	1	4	5	3,5	4	good
Karen (WSHiFM)	5	3	4	4	5	2	3	5	2	3	1	3,4	3	satisfactory
Wirtualny Doradca (PayU)	1	3	1	2	4	1	3	3	2	2	5	2,5	3	satisfactory
Ewa / Adam (Kredytum.pl)	5	4	4	4	3	3	3	5	2	1	1	3,2	3	satisfactory
Ewa (Orange)	5	5	5	5	5	3	5	5	2	5	5	4,5	5	very good

Source: own elaboration

Conclusions

This paper has presented a study of existing Polish-speaking commercial chatbots. The focus has been given on virtual assistants that work in the B2C sector, reach the widest possible range of users, and are presumably the most advanced commercial deployments of their creators. We evaluated ten quality components of every commercial chatbot, namely: visual look, form of implementation on the website, speech synthesis unit, built-in knowledge base (with general and specialized information), presentation of knowledge and additional functionalities, conversa-

tional abilities and context sensitiveness, personality traits, personalization options, emergency responses in unexpected situations, possibility of rating chatbot and the website by the user. By analyzing these factors we presented the current condition of Polish market of commercial virtual assistants and demonstrated the importance of performance, usability and overall quality evaluation of every commercial application of virtual assistant.

References

i. Abu Shawar B., Atwell E., Different measurement metrics to evaluate a chatbot system, [in:] Bridging the Gap: Academic and Industrial Research in Dialog Technologies, NAACL'07 Workshop Proceedings, Rochester, NY 2007 http://dx.doi.org/10.3115/1556328.1556341

ii. Berry D., Butler L., de Rosis F., Laaksolahti J., Pelachaud C., Steedman M., Embodied Believable Agents, MagiCster - Final evaluation report, University of Edinburgh, Edinburgh 2004

iii. Brahnam S., Strategies for handling customer abuse of ECAs, Proceedings of the INTERACT 2005 Workshop on Abuse: The darker side of Human-Computer Interaction, Rome 2005

iv. Brahnam S., Gendered Bods and Bot Abuse, Proceedings of the Workshop on Abuse and Misuse of Interactive Technologies, CHI 2006, Montreal 2006

v. Callejas Z., Griol D., López-Cózar R., Predicting user mental states in spoken dialogue systems, EURASIP Journal on Advances in Signal Processing, Vol. 6, 2011a http://dx.doi.org/10.1186/1687-6180-2011-6

vi. Callejas Z., López-Cózar R., Abalos N., Griol D., Affective Conversational Agents: the role of personality and emotion in spoken interactions, [in:] Perez-Marin D., Pascual-Nieto I. (eds.), Conversational Agents and Natural Language Interaction: Techniques and Effective Practices, IGI Global Publishers, Hershey, PA 2011b http://dx.doi.org/10.4018/978-1-60960-617-6.ch009

vii. Cassell J., Bickmore T., Billinghurst M., Campbell L., Chang K., Vilhjálmsson H., Yan H., Embodiment in Conversational Interfaces: Rea, ACM CHI'99 Conference Proceedings, Pittsburgh, PA 1999a http://dx.doi.org/10.1145/302979.303150

viii. Cassell J., Bickmore T., Campbell L., Vilhjálmsson H., Yan H., Conversation as a System Framework: Designing Embodied Conversational Agents, [in:] Cassell J. et al. (eds.), Embodied Conversational Agents, MIT Press, Cambridge, MA 1999b

ix. De Angeli A., On verbal abuse towards chatterbots, Proceedings of the Workshop on Abuse and Misuse of Interactive Technologies, CHI 2006, Montreal 2006

x. De Angeli A., Carpenter R., Stupid computer! Abuse and social identities, Proceedings of the INTERACT 2005 Workshop on Abuse: The darker side of Human-Computer Interaction, Rome 2005

xi. Elkins A.C., Derrick D.C., The Sound of Trust: Voice as a Measurement of Trust During Interactions with Embodied Conversational Agents, Group Decision and Negotiation, Vol. 22, Issue 5, Springer 2013 http://dx.doi.org/10.1007/s10726-012-9339-x

xii. Friedlein A., Maintaining and Evolving Successful Commercial Web Sites, Elsevier 2003 http://dx.doi.org/10.1016/b978-155860830-6/50022-8

xiii. Gaudiano P., Kater K., ALife-WebGuide: an intelligent user interface for Web site navigation, Proceedings of the 5th International Conference on Intelligent User Interfaces, IUI 2000, ACM, New York 2000 http://dx.doi.org/10.1145/325737.325800

xiv. Gerhard M., Evaluating Embodied Conversational Agents in Collaborative Virtual Environments, [in:] Ruttkay Z., André E., Johnson W.L., Pelachaud C. (eds.), Evaluating Embodied Conversational Agents, Dagstuhl Seminar Proceedings, IBFI, Schloss Dagstuhl, Germany 2006

xv. Haake M., Embodied Pedagogical Agents: From Visual Impact to Pedagogical Implications, doctoral thesis, Lund University, Sweden 2009

xvi. Heudin J.C., A Bio-inspired Nano-Agent Architecture for Intelligent Agents, [in:] Usmani Z. (ed.), Web Intelligence and Intelligent Agents, InTech, 2010 http://dx.doi.org/10.5772/8368

xvii. Heudin J.C., A schizophrenic approach for intelligent conversational agents, Proceedings of 3rd International Conference on Agents and Artificial Intelligence ICAART 2011, Vol. 2 - Agents, Rome 2011 http://dx.doi.org/10.5220/0003183902510256

xviii. Hsu Y.C., Affective Interfaces of Embodied Conversational Agents: studies of hardware and character interfaces, doctoral thesis, University of New South Wales 2011

xix. Kasap Z., Magnenat-Thalmann N., Intelligent virtual humans with autonomy and personality: State-of-the-art, Intelligent Decision Technologies, Vol. 1, No.1-2, IOS Press 2007 http://dx.doi.org/10.1007/978-3-540-79868-2_2

xx. Kenny P., Parsons T.D., Embodied Conversational Virtual Patients, [in:] Perez-Marin D., Pascual-Nieto I. (eds.), Conversational Agents and Natural Language Interaction: Techniques and Effective Practices, IGI Global Publishers, Hershey, PA 2011

xxi. Kenny P., Parsons T.D., Gratch J., Leuski A., Rizzo A.A., Virtual Patients for Clinical Therapist Skills Training, [in:] Pelachaud C., Martin J.C., André E., Chollet G., Karpouzis K., Pelé D. (eds.), Intelligent Virtual Agents, LNCS, Vol. 4722, Springer, Heidelberg 2007 http://dx.doi.org/10.1007/978-3-540-74997-4_19

xxii. Kowalski S., Pavlovska K., Goldstein M., Two Case Studies in Using Chatbots for Security Training, [in:] Dodge R.C., Futcher L. (eds.), Information Assurance and Security Education and Training, IFIP Advances in Information and Communication Technology, Vol. 406, 2013 http://dx.doi.org/10.1007/978-3-642-39377-8_31

xxiii. Knoppel F., Tigelaar A., Oude Bos D., Alofs T., Ruttkay Z., Trackside DEIRA: A Dynamic Engaging Intelligent Reporter Agent, Proceedings of 7th International Conference on Autonomous Agents and Multiagent Systems, AAMAS08, Estoril, Portugal, IFAAMAS 2008

xxiv. Kopp S., Bergmann K., Buschmeier H., Sadeghipour A., Requirements and building blocks for sociable embodied agents, Proceedings of the 32nd annual German conference on Advances in artificial intelligence, Springer 2009 http://dx.doi.org/10.1007/978-3-642-04617-9_64

xxv. Kumar R., Rosé C.P., Building Conversational Agents with Basilica, NAACL-HLT 2009 Proceedings, Boulder, Colorado 2009 http://dx.doi.org/10.3115/1620959.1620961

xxvi. Loyall B., Bates J., Personality-Rich Believable Agents That Use Language, Proceedings of the First International Conference on Autonomous Agents, ACM, California 1997 http://dx.doi.org/10.1145/267658.267681

xxvii. Millet P., Heudin J.C., Web mining in the EVA intelligent agent architecture, Web Intelligence and Intelligent Agent Technology Workshops, ACM-WIC-IEEE 2007, California 2007 http://dx.doi.org/10.1109/wi-iatw.2007.76

xxviii. Ortony A., On making believable emotional agents believable, [in:] Trappl R., Petta P., Payr S. (eds.), Emotions in humans and artifacts, MIT Press, Cambridge, MA 2003

xxix. Sjödén B., Silvervarg A., Haake M., Gulz A., Extending an Educational Math Game with a Pedagogical Conversational Agent: Facing Design Challenges, Communications in Computer and Information Science, No. 126, Springer Verlag 2011 http://dx.doi.org/10.1007/978-3-642-20074-8_10

xxx. Stavropoulou P., Spiliotopoulos D., Kouroupetroglou G., Design and Development of an Automated Voice Agent: Theory and Practice Brought Together, [in:] Perez-Marin D.,

Pascual-Nieto I. (eds.), Conversational Agents and Natural Language Interaction: Techniques and Effective Practices, IGI Global Publishers, Hershey, PA 2011 http://dx.doi.org/10.4018/978-1-60960-617-6.ch015

xxxi. Stern A., Creating emotional relationships with virtual characters, [in:] Trappl R., Petta P., Payr S. (eds.), Emotions in humans and artifacts, MIT Press, Cambridge, MA 2003

xxxii. Thomas C., Fischer G., Using Agents to Personalize the Web, IUI'97 Proceedings of the 2nd international conference on Intelligent user interfaces, ACM, New York 1997 http://dx.doi.org/10.1145/238218.238287

xxxiii. Traum D., Swartout W., Gratch J., Marsella S., A virtual human dialogue model for non-team interaction, [in:] Dybkjaer L., Minker W. (eds.), Recent Trends in Discourse and Dialogue, Antwerp, Belgium, Springer 2008 http://dx.doi.org/10.1007/978-1-4020-6821-8_3

xxxiv. Van Deemter K., Krenn B., Piwek P., Klesen M., Schroeder M., Baumann S., Fully Generated Scripted Dialogue for Embodied Agents, Artificial Intelligence Journal, Vol. 172, Issue 10, 2008 http://dx.doi.org/10.1016/j.artint.2008.02.002

xxxv. Van Vugt H., Bailenson J., Hoorn J., Konijn E., Effects of facial similarity on user responses to embodied agents, ACM Transactions on Computer-Human Interaction, Vol. 17, No. 2, Article 7, New York 2010 http://dx.doi.org/10.1145/1746259.1746261

xxxvi. Xiao J., Stasko J., Catrambone R., The Role of Choice and Customization on Users' Interaction with Embodied Conversational Agents: Effects on Perception and Performance, Proceedings of CHI 2007, San Jose, CA 2007 http://dx.doi.org/10.1145/1240624.1240820