ACTA UNIVERSITATIS AGRICULTURAE ET SILVICULTURAE MENDELIANAE BRUNENSIS

Volume 67 138 Number 6, 2019

CHATBOTS FOR ENTERPRISES: OUTLOOK

Pavel Kostelník[1], Ivo Pisařovic[1], Mikuláš Muroň[1], František Dařena[1], David Procházka[1]

[1] Department of Informatics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

To link to this article: https://doi.org/10.11118/actaun201967061541
Received: 3. 6. 2019, Accepted: 5. 11. 2019

To cite this article: KOSTELNÍK PAVEL, PISAŘOVIC IVO, MUROŇ MIKULÁŠ, DAŘENA FRANTIŠEK, PROCHÁZKA DAVID. 2019. Chatbots for Enterprises: Outlook. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis,* 67(6): 1541–1550.

Abstract

Chatbots are going to be the main tool for automated conversations with customers. Still, there is no consistent methodology for choosing a suitable chatbot platform for a particular business. This paper proposes a new method for chatbot platform evaluation. To describe the current state of chatbot platforms, two high-level approaches to chatbot platform design are discussed and compared. WYSIWYG platforms aim to simplicity but may lack some advanced features. All-purpose chatbot platforms require extensive technical skills and are more expensive but give their users more freedom in chatbot design. We provide an evaluation of six major chatbot solutions. The proposed method for the chatbot selection is demonstrated on two sample businesses – a large bank and a small taxi service.

Keywords: chatbot, conversational agent, natural language processing, conversational commerce, intelligent assistant, customer relationship management, chatbot market overview

INTRODUCTION

Generally, chatbots are one of the means of human-computer interaction. They provide a new type of user interface (conversational interface) as a service. This interaction is designed as a dialog/conversation between a user and a chatbot.

Schumaker *et al.* (2007) define a chatbot as a system that "seeks to mimic conversation rather than understand it". This is a more modern definition than the previous definition by Mauldin (1994), who defines chatbot as a system that aims at passing the Turing Test as defined in Turing (1950). Mauldin (1994) used term Chatter-Bot as the name for any dialog system capable of having informal conversations.

One of the first systems which can be considered a chatbot is ELIZA developed by Joseph Weizenbaum in 1966 at MIT. ELIZA was able to mimic human-like responses in a conversation by rephrasing input sentences (Weizenbaum, 1966). This system used rule-based logic and was therefore very limited. However, it set the path for future development. Since then, technology has improved a lot. This is proven by current successful systems, such as Mitsuku made by Steve Worswick, which has won prestigious Loebner prize multiple times and was rated in the user's study better than the widely used commercial systems (Worswick, 2019).

Chatbots are becoming popular in a wide variety of business applications. They prove to be useful in marketing, consumer support and FAQ. They can also provide support for sales representatives or even perform sales themselves (Shawar and Atwell, 2007).

Chatbots are also a great tool for productivity according to Brandtzaeg and Følstad (2017). They studied main motivational factors driving chatbot use and category "productivity" was among the most reported uses. Other factors driving chatbots use are social, relational and also entertainment.

Emotional engagement with users is also very important. Humans tend to expect some emotional impact from a conversation. Xu *et al.* (2017) study

on customer service chatbots shows that 40% of user requests are in their nature emotional. Thus, chatbot's ability to provide emotional relatability is an important factor for overall customer satisfaction.

Chatbots can be utilized in many different areas of life, one of the common examples being primary care. The main challenge in this field represents the patient intake process which can be very labor-intensive. Parts of this labor can be simplified by automating it using chatbots. Ni *et al.* (2017) developed a system called Mandy specially focused on this task. This system consists of a mobile app with a chatbot for the patients and a diagnostic unit. The key task is to understand the patient's symptoms and map them to potential diseases. In this task, Mandy achieves prediction accuracy 67% (proportion of correctly matched hypotheses).

Other successful examples of chatbot system deployment can be found in the e-commerce domain. Cui *et al.* (2017) present a system for providing information about products called SuperAgent. This system is able to answer questions about various technical specifications of products (e.g. what is the screen resolution?) and also provide a summary of customer reviews. Contrary to systems with similar task SuperAgent leverages existing product descriptions and user-generated content from various e-commerce websites, hence the time spent by building knowledge base is significantly reduced. Usability analysis done by authors shows that SuperAgent has improved end-to-end user experience for particular e-commerce websites.

Heo and Lee (2018) present a single-case study of Naver TalkTalk. This case study is focused on Naver Inc, where TalkTalk chatbot was developed as a tool to support online shopping. After experimentation Naver company realized that chatbots do not work as expected if they only rely on NLU. They have achieved better results by implementing a service that can provide users with scenarios and let users choose among them (a card-based chatbot). By adopting this approach company's compliance rate were up by 29% and purchase conversion up by 12.4% in the first month (Heo nad Lee, 2018).

Existing Work

Braun and Matthes (2019) propose a framework to classify chatbots based on selected important features. These features belong to 5 main topics (I/O, timing, flow, platform, understanding). Criteria selected by Braun and Matthes (2019) are focused on evaluating functionality. This approach is very useful in general settings, but does not provide easy to use decision mechanism for business-focused applications.

Braun *et al.* (2017) are using a similar approach in comparison of Natural Language Understanding services. NLU as service plays an important role in the recent popularity of chatbots and comparison of these services is beneficial. Different NLU services (Watson, API.ai, wit.ai, etc.) are evaluated in three main aspects: intents, entities and batch import. Comparison criteria presented in this research are helpful in the decision-making process but may be too technical for business-oriented decision making.

Kuligowska (2015) explores the potential of commercial applications of chatbots and mapping their success in real-world B2C deployment. Ten various aspects of chatbots were selected (knowledge base, conversational abilities, language skills, context sensitiveness etc.) and subjected to comprehensive analysis. Evaluation of these aspects is used as metrics to estimate the overall quality of selected chatbots. Contrary to our approach focused on platform comparison, Kuligowska (2015) focuses on the comparison of particular chatbot deployments.

Hill *et al.* (2015) focus on an analysis of differences between human-chatbot conversation and human-human conversation. This study analyzed seven variables (words per conversation, messages per conversation, average number of words per message, etc.) from real human-human and human-chatbot conversations. Among other findings, it states that people send more than twice as many messages to chatbots compared to a human-human conversation.

Most of the current literature focuses on evaluating the impact of chatbots in specific case studies or on various aspects of chatbots in the field of human-computer interaction. Selection of the right platform for specific business use case can be challenging and currently available literature does not provide any comprehensive methodology.

The main contribution of this paper is to present a method for selecting a chatbot platform based on characteristics of individual businesses. This method is demonstrated on two sample businesses that represent marginal companies on the market – a small business and a large enterprise. Different chatbot platforms focus on different kind of businesses. There is no universal platform for any kind of business as is described in the following chapter.

Review of the Existing Technical Solutions

A number of platforms and packaged solutions are becoming available for chatbot builders to speed up development and deployment of chatbots. Usage of high-level tools does not require many technical skills. Various categorizations are used in published works but many of them involve evaluation of framework ease of use (Srivastava and Prabhakar, 2019; Galitsky, 2019). These criteria divide platforms into two major groups – we called them WYSIWYG and all-purpose chatbot platforms.

Before describing individual platforms, we must define basic terms used in the following texts.

INTENT is something a user wants to know or get, for example “display current weather”. ENTITY is one or a few words used to parametrize the intent. E.g. “Brno” can be an entity in an intent “display current weather in {city}”. WEBHOOK is used to perform actions on external servers that the chatbot cannot do itself, e.g. save an order made in the chat to a company information system (Shevat, 2017).

WYSIWYG chatbot platforms – In other sources also mentioned as high-level chatbot platforms or trivial-bot platforms (Galitsky, 2019; Rothel, 2019). These implementations tie very closely to actual business representatives who recognize the opportunity of chatbot integration in their current or potential market (Singh *et al*., 2019). This category includes platforms such as Chatfuel or ManyChat.

Target customers of these platforms are marketers, salesmen or public relations associates. These workers lack the ability to program but they bring valuable knowledge of customers.

These chatbot platforms aim to greatly simplify chatbot development and bring down the technical barriers that come along. Their focus is on delivering core added values of chatbots fast and easily to contribute to competitive advantage (Waghmare, 2019). Many tuning aspects are set up using the good defaults design patterns and the platform usually uses pre-trained datasets and simple keyword matching for extracting intents (Janarthanam, 2017). WYSIWYG chatbot platforms tend to be purely cloud-based with a strong focus on a very understandable and friendly user interface.

All-purpose chatbot platforms – All-purpose chatbot platforms give their users full freedom to combine and integrate their chatbot with tools of their choice (Singh *et al*., 2019). Dialogflow, IBM Watson Conversation and Microsoft Bot Framework are a few examples of services in this category. Other sources refer to this group of platforms as conversational or conversational assistant platforms (Rothel, 2019; Akhtar *et al*., 2019)

Target customers are expected to have a certain level of technical skills that are required to leverage the platform. These platforms usually offer very rich comprehensive APIs, pre-built clients for multiple programming languages and give their users real freedom in implementation (Galitsky, 2019). Multiple services and applications can be integrated and thus deliver a wide set of functionalities to a chatbot. This freedom is weighted out by more complex setup and requires programming skills.

The following text discusses the most well-known and adopted platforms with descriptions of their specifics and advantages or disadvantages. Particular platform selection is compiled on multiple sources. The aim is to represent typical approaches to chatbot creation. Evaluated solutions do not include platforms that lack the full-stack support (building chatbot software from conversation design to end-user delivery and testing).

Chatfuel

Chatfuel is the most adopted platform by far in terms of user traffic. It has been used to create over 360,000 chatbots serving more than 17 million users (Root Info Solutions, 2018). Chatfuel aims at marketers, campaign promotions, and simplicity.

Bots built using Chatfuel tend to be one-purpose only and serve that use case well. This platform features a web conversation builder that is easy to understand and helps to visualize the flow of the conversation. Chatbots built using this tool are usually narrow-guided and do not offer a lot of freedom. Questions are direct and one block of conversation leads directly to another without allowances for a deviation.

The builder, however, supports various conversation elements including components like an image gallery or calling an external webhook. Chatfuel also supports broadcasting and has basic reporting and analytic tools.

The platform strongly focuses on Facebook. This platform gains its adoption mostly thanks to the already existing strong base of users in Facebook groups and pages.

ManyChat

ManyChat is similar to Chatfuel in its concept. It tries to abstract bot building into a form of a conversational designer and supportive advertisement tool. To simplify chatbot development even further, it offers a set of predefined templates that can be adjusted for a specific business.

ManyChat also offers a wide range of social marketing tools like the audience growth potential analytics, messaging platform promotions, or message broadcasting. ManyChat is strictly tied to Facebook Messenger as a delivery channel platform. The option to call external webhooks has been introduced in early 2018 and is available only in a paid version. ManyChat also offers an option for chatbot owners to overtake the conversation.

Features that ManyChat and Chatfuel offer are similar. While Chatfuel lacks some features (promotions, live chat, some conversation builder components), it feels much more like a unified platform (Johnson, 2018).

Botsify

Botsify is a platform that stands between the traditional WYSIWYG-focused chatbot platforms and all-purpose chatbot platforms. Botsify includes a drag-and-drop conversational designer similar to those present in the Chatfuel or ManyChat platforms. Additionally, it also offers an integrated AI and automated entity recognition.

The goal is to keep the ownership of chatbots within the business segment of organizations (Nath, 2018). Not only Botsify offers endpoint channel integration to Facebook Messenger, but also includes a webchat (limited to WordPress only) support and is deployable to Amazon Alexa.

Microsoft Bot Framework

Microsoft Bot Framework is a very wide set of tools that allows full-stack bot creation. It is an all-purpose solution requiring IT experts for deployment.

The core lies in Microsoft Bot Builder SDK. This set of libraries is natively available for implementation using C# or Node.js. The library itself is focused on building a core conversation interface for a bot using one of these languages (others can leverage on REST API tied to Azure Bot Service). The SDK is a bare-bone foundation for conversation design (no UI conversation designer is included) and UI components definition.

Language understanding is completely abstracted from the bot builder SDK. Microsoft offers a product called LUIS (Language Understanding Intelligent Service) that provides services like intent detection or entity recognition. Microsoft Bot Framework also includes Azure Cognitive Services that can enhance bot functionality (Machiraju and Modi, 2018).

The whole framework supports on-premise deployment as well as Azure cloud deployment (some calls, e.g. LUIS API, are permanently bound to Azure). Microsoft Bot Framework allows delivery to many messaging channels including Telegram, Slack, a standalone website, Facebook Messenger, Skype, or Outlook e-mail.

Microsoft Bot Framework is a great example of a complex all-purpose chatbot platform. It has many parts that need to be connected manually and set up to accommodate for a final functional application. At the same, time the freedom that Microsoft platform brings is incomparable to others.

Dialogflow

This all-purpose platform offers one of the top-class NLP intent and entity recognition (Canonico and Russis, 2018) including supervised training and manual results correction.

Dialogflow is a platform focused on core NLP. Intents and conversations can be defined (using stateful contexts) but the platform does not feature any rich graphical tool. Yet it is simple enough that basic transactional chatbots can be created with minimal effort. According to Faatz (2018), transactional chatbots are characterized as bots limited to a number of optimized use cases that can eliminate the need to talk to an expert or use complex UI.

A webhook interaction is available via a direct call to an external system or using Google Cloud Functions. Dialogflow supports the following channels for message delivery: Slack, Twilio, Viber, Skype, Telegram, Facebook Messenger, Amazon Alexa, Cortana, and, of course, Google Assistant. Dialogflow can be fully controlled via a rich REST client.

IBM Watson Conversation

IBM has heavily invested into NLP and machine learning. Today IBM provides Watson Conversation Service as its flagship for NLU.

This platform is a cloud-only based platform running on IBM Cloud. Watson Conversations includes a high-level UI conversation designer, entity recognition, and a basic set of analytic tools. The IBM platform behaves like a typical cloud SaaS. The Watson engine can serve as an independent module that only provides NLP services (Biwas, 2018).

It is similar to Dialogflow with the exception that its usage concept is reversed. Dialogflow calls external service webhook as a part of a response generation, while IBM Watson Conversations only exposes its APIs and relies on an external application orchestrating and delegating user queries.

Evaluating Chatbot Platforms

An overview of existing technical solutions for chatbot deployment in a business environment was presented. A set of criteria that will be used to evaluate particular platforms is to be defined.

The goal of the evaluation presented through the means of this paper is the focus on the business application. In particular, selecting the best chatbot platform for a business to implement a chatbot solution to improve customer experience. Therefore, it is necessary to make a clear distinction between the goals and aims that various businesses have with their chatbot. We consider company size the most important division factor for this evaluation. Other key distinctive factors could include budget, customer base characteristics or business model. For demonstration, we consider two sample companies and their use cases:

1. Local taxi company – A small company of 50 employees. All IT demands are outsourced. The company marketing manager plans to deploy a chatbot as an alternative way to order taxi services. Phone calls can be problematic and sometimes the customer prefers a different kind of interaction than a phone call. Calls are also a subject to human error. Ordering a taxi via a web form is not comfortable for the customers. Therefore, a chatbot would provide a comfortable and reliable way to place orders of services. Low costs and simplicity are the main requirements for the chatbot.
2. Bank – A large company having 10,000 employees providing financial services at the national level. This company has its own IT department. The primary goal of a chatbot deployment is to reduce costs of helpdesk. Such a company is willing to invest a high amount of money into the chatbot development. A group within the IT department coordinates with the management. A strategy must be developed to integrate chatbot services into other company systems and to support business goals of the

company. Data security is also an important concern for this use case.

Evaluation method discussed below reflects differences in various companies for the purpose of this article. The advised method aims to be universal for any use case with appropriate adjustments of weights as described below.

The approach of choosing use case for a smaller business and a larger enterprise to evaluate the subject matter underlines different needs and expectations for these two subjects (Ayu *et al.*, 2019) and mimics methodology mentioned in other works (Nogues and Valladares, 2017).

Criteria for Platform Evaluation

A set of criteria needs to be established to compare a wide variety of functionality that a business application of a chatbot requires. Criteria selection is complex matter depending on multiple factors (Braun and Matthes, 2019). The following criteria represent a comprehensive group of different platform characteristics that are measurable and specifiable. The criteria were derived from references and characteristics of the examined chatbot platforms, as described earlier.

- **NLP maturity** – Communication with a chatbot uses natural language and intent has to be extracted from user input. Thus, processing of natural language is essential for adequate chatbot response. We practically tested it by providing the same input to individual platforms and comparing the correctness of an inferred intent.

  Some platforms are leaning toward a straightforward approach where dialogue flow strictly leads from one intent to another based on a direct user reply. Other solutions allow dynamic intent detection at any moment during a conversation flow based on an NLP analysis. Optionally, the current session state can provide an additional context if supported by the platform.
- **Additional capabilities** – NLP is essential for chatbot operation, but other related services can bring additional functionality and simplify chatbot development. E.g. a bot designed for tracking personal expenses should detect the expenses in the images of receipts uploaded to the conversation thanks to built-in image recognition service. Some platforms offer such services seamlessly, while others lack them completely and force manual implementation.
- **API interaction** – Support of webhooks is usually essential if a chatbot is to bring an added value to a bigger enterprise.

  Services that do not offer webhooks are limited to data provided directly in a chatbot definition and lead towards data structure representations embedded in the chatbot itself. In such cases, intents have to fulfil user's needs by providing plain information only in a form of messages that are bound to the intent detection logic (no additional queries to external systems are available).

  On the top of webhooks, additional API integrations are very valuable since they enable connection to already existing systems and databases. Events in other systems can serve as a trigger for chatbot interaction – e.g., whenever status of an order changes in a warehouse management system, a related chatbot can inform the user via a message stating that that the status of the order has changed.
- **Delivery channels support** – Chatbot services need to plug in their chatbots to an existing delivery channel (webchat, Facebook Messenger, Slack, Amazon Alexa, or others). These channels already house a very strong user base that can take advantage of the services offered by the chatbot.
- **Analytic services** – Some platforms include analytics and feedback services that can help to improve the chatbot design. Such analytics can detect a potential growth opportunity, incorrectly recognized intents, and other defects in a conversation flow.
- **Vendor lock-in** – Transferring intents or detected user entities across platforms usually represents an added value. This gives developers freedom to move away from a particular implementation without having to rewrite a chatbot from scratch.
- **Auto-configuration** – Some platforms aim for maximum auto-configuration while others do not provide this setup that tends to limit the solution. Auto-configuration and focus on business ownership is especially important for organizations that have small or no IT group that will maintain the chatbot.
- **Price** – This attribute characterizes the price of the platform and related components. When the platform is provided on the basis of a subscription, the cost can be estimated according to the cost of related service (Azure cloud in the case of Microsoft Bot Framework and Google Cloud in the case of Dialogflow).

MATERIALS AND METHODS

Evaluation Method

AHP is a frequently used tool in decision making (Vaidya and Kumar, 2006) and is used in many fields related to information technologies (Salmeron and Herrero, 2005). This method is suitable for the purpose of this paper because it is easily understandable by general audience and its results are transparent. It also allows to measure non-quantifiable criteria.

Firstly, features of individual chatbot platforms described in chapter Review of the Existing Technical Solutions were scored using the scale in Tab. I.

Afterwards, the weights of the criteria were calculated. The quantification of preferences

I: *Scores for evaluating individual features of chatbot platforms*

Score	Description
1	Support for this feature is unavailable or this feature does not apply.
2	Some or very limited support for this feature is available thanks to e.g. workarounds or other third-party software.
3	The feature is available, but polishing is required. Has certain limitations that affect its functionality or usability.
4	The feature is available in the system, lacks some minor details like complete documentation or advanced functionality.
5	Support for this feature in the system is fully present and has no limitations.

regarding two criteria A and B from the criteria hierarchy can be found in Tab. II. The final weight is calculated as geometric mean of the quantified relative importance of the criteria as is shown in Eq. 1 (Saaty, 1980). Finally, the final weighted score for each platform and use case was calculated using dot product, see Eq. 2.

$$g_c = \left(\prod_{i=1}^{n} x_i \right)^{\frac{1}{n}} = \sqrt[n]{x_1 x_2 \ldots x_n}$$

$$w_c = \frac{g_c}{\sum_{j=1}^{n} g_j} \tag{1}$$

Geometric mean g_c was used to make an average relative importance of a criterion c to other criteria. The criterion weight w_c was calculated as the geometric mean divided by the sum of geometric mean of all criteria (normalization).

$$\vec{w}_b \cdot \vec{s}_p = \sum_{c=1}^{n} w_c s_c \tag{2}$$

II: *Weights for criteria comparison using Saaty method*

Weight	Description
1	Both criteria are equal
3	Weakly preferred
5	Strongly preferred
7	Very strongly preferred
9	Absolutely preferred

Final weighted score for each platform p and business b was calculated as dot product of criteria weights w and scores s.

RESULTS AND DISCUSSION

Individual chatbot platforms were evaluated based on the score scale described in the chapter Evaluation method. Each criterion was given two weights based on the requirements of our two use cases (see Tab. III and Tab. IV). The results for both use cases are calculated as dot product of weights and scores according to the previous chapter. The complete evaluation results are shown in Tab. V.

III: *Calculating the weights of criteria for the small taxi company*

	NLP maturity	Additional services	API interaction	Delivery Channels	Analytics	Not vendor lock-in	Auto configuration	Price	Geometric mean	Weights
NLP maturity	1.00	3.00	0.20	3.00	0.20	1.00	0.11	0.11	0.51	0.03
Additional services	0.33	1.00	0.11	0.11	0.20	0.11	0.11	0.11	0.18	0.01
API interaction	5.00	9.00	1.00	5.00	1.00	5.00	0.11	0.11	1.39	0.09
Delivery Channels	0.33	9.00	0.20	1.00	1.00	3.00	0.11	0.11	0.62	0.04
Analytics	5.00	5.00	1.00	1.00	1.00	3.00	0.11	0.11	0.99	0.07
Not vendor locked-in	1.00	9.00	0.20	0.33	0.33	1.00	0.11	0.11	0.47	0.03
Auto configuration	9.00	9.00	9.00	9.00	9.00	9.00	1.00	9.00	6.84	0.46
Price	9.00	9.00	9.00	9.00	9.00	9.00	0.11	1.00	3.95	0.26
									14.95	1.00

When comparing individual criteria, characteristics of sample use case companies were taken into account. E.g. the small company (represented by the local taxi service) counts auto-configuration and price among the most important criteria and therefore they were assigned a higher number. Analytics or vendor lock-in does not represent a big concern thus their significance is minimized in this use case. Tab. III depicts the final normalized weights for the small company. It reflects the need to deploy a chatbot without extra costs or special functions (in this use case a simple solution is preferred).

As seen in Tab. IV, for the second use case (the large enterprise), API Interaction is the most important criteria. NLP Maturity, Delivery channels, Analytics and Vendor lock-in are less important. This demonstrates the needs of a larger company – the company requires advanced functions to use the full potential of the chatbot.

Based on the provided evaluation for the small taxi company, Chatfuel is the best platform for this use case. Chatfuel is strongly focused on auto-configuration, tied to Facebook as the delivery channel, and provides good analytics. These facts make it a perfect choice in a use case trying to reasonably limit expenses for chatbot creation. Microsoft Bot Framework was evaluated as the worst platform choice for the small company. It requires technical skills and high costs and as such is not suitable for the small local company

IV: *Calculating the weights of criteria for the large bank*

	NLP maturity	Additional services	API interaction	Delivery Channels	Analytics	Not vendor locked-in	Auto configuration	Price	Geom. mean	Weights
NLP maturity	1.00	5.00	0.11	3.00	0.33	1.00	5.00	3.00	1.30	0.11
Additional services	0.20	1.00	0.14	0.20	0.14	0.14	1.00	1.00	0.32	0.03
API interaction	9.00	7.00	1.00	7.00	1.00	5.00	9.00	9.00	4.53	0.39
Delivery Channels	0.33	5.00	0.14	1.00	3.00	1.00	5.00	3.00	1.35	0.12
Analytics	3.00	7.00	1.00	0.33	1.00	3.00	5.00	3.00	2.05	0.18
Not vendor locked-in	1.00	7.00	0.20	1.00	0.33	1.00	7.00	3.00	1.33	0.11
Auto configuration	0.20	1.00	0.11	0.20	0.20	0.14	1.00	0.20	0.27	0.02
Price	0.33	1.00	0.11	0.33	0.33	0.33	5.00	1.00	0.54	0.05
									11.69	**1.00**

V: *Evaluation of the chatbot platforms*

	Weights for the taxi company	Weights for the bank	WYSIWYG-oriented platforms			All-purpose platforms		
			Chatfuel	ManyChat	Botsify	Microsoft Bot Framework	Dialogflow	IBM Watson
NLP maturity	0.03	0.11	3.00	2.00	2.00	5.00	5.00	5.00
Additional services	0.01	0.03	1.00	1.00	1.00	5.00	3.00	5.00
API interaction	0.09	0.39	4.00	4.00	3.00	5.00	5.00	2.00
Delivery Channels	0.04	0.12	3.00	2.00	3.00	4.00	5.00	2.00
Analytics	0.07	0.18	4.00	3.00	1.00	2.00	5.00	2.00
Not vendor locked-in	0.03	0.11	3.00	2.00	2.00	1.00	4.00	1.00
Auto configuration	0.46	0.02	5.00	5.00	4.00	1.00	1.00	2.00
Price	0.26	0.05	3.00	3.00	4.00	1.00	1.00	1.00
Weighted Total (small business)			**4.05**	**3.88**	**3.50**	**1.75**	**2.06**	**1.84**
Weighted Total (large business)			**3.55**	**3.04**	**2.44**	**3.63**	**4.56**	**2.26**

without an IT department. It can also be concluded that all the WYSIWYG platforms are definitely relevant for the small company use case.

Results show Dialogflow platform as the best solution for the large company (the national bank as mentioned in the second use case). Choosing Dialogflow is very reasonable thanks to rich API options and the overall high quality of NLP. The use case describes a company with a dedicated IT team, that can leverage the full potential of the platform. All other platforms have a significantly lower score. That puts Dialogflow as the leader for use in larger companies. The worst ranking platform for this use case is IBM Watson because this platform does not offer the full range of features and is itself mainly considered just as an additional service.

Various businesses have very different aims with the chatbot application, and this paper tries to indicate that by choosing two particular use cases from opposite sides of the spectrum. Chatbot applications deployments exist in both categories – smaller local companies and larger national or international businesses (Waghmare, 2019) – thus the reason to represent both separately in this review.

An obvious limitation of this evaluation lies in the limited number of features evaluated or appropriate specifics bound to individual business segments. This paper suggests a generic guideline on evaluation, a list of important characteristics of chatbot platforms and how to apply a simple method for particular platform selection in a business use case.

CONCLUSION

The goal of this paper was to introduce the current state of commercial chatbot platforms and propose an evaluation method for chatbot selection according to needs of different companies.

The first part of the paper describes the chatbot definition and use of chatbots in different organizations. Based on the literature review, it was concluded there is no universal business-oriented methodology for selecting a chatbot platform for a particular business.

Secondly, six popular chatbot platforms are described, three WYSIWYG and three all-purpose platforms. WYSIWYG platforms are more user-friendly and enable to build a simple chatbot more easily. All-purpose platforms have a lot of freedom but require technical skills and are more costly.

Finally, we proposed a method based on AHP. The suitability of the reviewed platforms for businesses of different sizes has been evaluated using this method. We defined two sample companies (the taxi company and the bank) and set up relevant weights. The results show that Chatfuel is the best choice for the small taxi company and Dialogflow is the best choice for the large bank.

Thanks to the proposed method of evaluation, any company can determine the weights according to its specific requirements and select a chatbot platform that fits best its customers and business strategy. In comparison to other work like Braun *et al.* (2017), our paper is focused on particular business requirements, not only a comparison of technical features.

REFERENCES

AKHTAR, M., NEIDHARDT, J. and WERTHNER, H. 2019. The Potential of Chatbots: Analysis of Chatbot Conversations. In: *2019 IEEE 21st Conference on Business Informatics (CBI)*. 15–17 July 2019, Moscow, Russia. IEEE, pp. 397–404.

AYU, P. D., IDA, R. and APIT, S. W. 2019. The Influence of Company Size and Audit Fee on Audit Quality. Proceedings of the 1st International Conference on Economics, Business, Entrepreneurship, and Finance (ICEBEF 2018). *Advances in Economics, Business and Management Research*, 65: 371–375.

BIWAS, M. 2018. *Beginning AI Bot Frameworks: getting started with bot development*. 1st Edition. New York: Apress.

BRANDTZAEG, P. B. and FØLSTAD, A. 2017. Why people use chatbots. In: *International Conference on Internet Science.* Cham: Springer, pp. 377–392.

BRAUN, D., HERNANDEZ-MENDEZ, A., MATTHES, F. and LANGEN, M. 2017. Evaluating natural language understanding services for conversational question answering systems. In: *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue.* August 2017, Saarbrücken, Germany. ACL, pp. 174–185.

BRAUN, D. and MATTHES, F. 2019. Towards a Framework for Classifying Chatbots. Proceedings of the International Conference on Enterprise Information Systems (ICEIS). *Advances in Economics, Business and Management Research*, 65(1/2): 496–501.

CANONICO, M. and RUSSIS, L. 2018. A Comparison and Critique of Natural Language Understanding Tools. In: *The Ninth International Conference on Cloud Computing, GRIDs, and Virtualization*. IARIA, pp. 110–115.

CUI, L., HUANG, S., WEI, F., TAN, C., DUAN, C. and ZHOU, M. 2017, SuperAgent: A customer service chatbot for e-commerce websites. In: *Proceedings of ACL 2017 System Demonstrations*. Vancouver, Canada, July 30 – August 4, 2017. Association for Computational Linguistic, pp. 97–102.

FAATZ, S. 2018. *Transaction or Knowledge Chatbot? Moving Beyond the Siri Syndrome.* [Online]. Available at: https://www.telerik.com/blogs/transactional-or-knowledge-chatbot-moving-beyond-siri-syndrome [Accessed: 2019, September 9].

GALITSKY, B. 2019. *Developing Enterprise Chatbots: Learning Linguistic Structures*. 1st Edition. Zurich: Springer International Publishing.

HEO, M. and LEE, K. J. 2018. Chatbot as a new business communication tool: The case of naver talktalk. *Business Communication Research and Practice*, 1(1): 41–45.

HILL, J., FORD, W. R. and FARRERAS, I. G. 2015. Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. *Computers in Human Behavior*, 49: 245–250.

JANARTHANAM, S. 2017. *Hands-on chatbots and conversational UI development: build chatbots and voice user interfaces with Chatfuel, Dialogflow, Microsoft Bot Framework, Twilio, and Alexa Skills*. Birmingham: Packt Publishing.

JOHNSON, M. 2018. *Why You Should NOT Build a ChatBot on Manychat or Chatfuel and What to do Instead.* [Online]. Available at: https://messengerfunnels.com/why-you-should-not-build-a-chatbot-on-manychat-or-chatfuel-and-what-to-do-instead/ [Accessed: 02/14/2019].

KULIGOWSKA, K. 2015. Commercial chatbot: Performance evaluation, usability metrics and quality standards of embodied conversational agents. *Professionals Center for Business Research*, 2(2): 22.

MACHIRAJU, S. and MODI, R. 2018. *Developing Bots with Microsoft Bots Framework: Create Intelligent Bots using MS Bot Framework and Azure Cognitive Services*. 1st Edition. New York: Apress.

MAULDIN, M. 1994. Chatterbots, Tinymuds, and the Turing Test: Entering the Loebner Prize Competition. In: *AAAI 94 Proceedings.* Seattle: AAAI, pp. 16–21.

NATH, J. 2018. How to Identify the Right Platform for Your Chatbot. *Chatbots Magazine*. [Online]. Available at: https://chatbotsmagazine.com/identifying-the-platform-for-your-chatbot-312e46ef9e8a [Accessed: 2019, March 20].

NI, L., LU, C., LIU, N. and LIU, J. 2017. Mandy: Towards a smart primary care chatbot application. In: *International Symposium on Knowledge and Systems Sciences*. Singapore: Springer, pp. 38–52.

NOGUES, A. and VALLANDERES, J. 2017. *Business Intelligence Tools for Small Companies A Guide to Free and Low-Cost Solutions*. 1st Edition. New York: Apress.

ROOT INFO SOLUTIONS. 2018. *Chatbot Development With Chatfuel Platform*. [Online]. Available at: https://www.rootinfosol.com/chatbot-development-chatfuel-platform [Accessed: 2019, March 7].

ROTHEL, I. 2019. *How to choose the right chatbot platform*. [Online]. Available at: https://www.predictiveanalyticstoday.com/what-is-chatbot-platform/#howtochoosetherightchatbotplatforms [Accessed: 2019, October 14].

SAATY, T. L. 1980. *The Analytic Hierarchy Process.* New York: McGraw Hill.

SALMERON, J. and HERRERO, I. 2005. An AHP-based methodology to rank critical success factors of executive information systems. *Computer Standards & Interfaces*, 28(1): 1–12.

SCHUMAKER, R., GINSBURG, M. *et al.* 2007. An evaluation of the chat and knowledge delivery components of a low-level dialog system: The AZ-ALICE experiment. *Decision Support Systems*, 42(4): 2236–2246.

SINGH, A., RAMASUBRAMANIAN, K. *et al.* 2019. *Building an Enterprise Chatbot*. New York: Apress.

SHAWAR A. B. and ATWELL, E. 2007. Chatbots: Are they Really Useful? *LDV Forum*, 22(1): 29–49.

SHEVAT, A. 2017. *Designing bots: creating conversational experiences*. 1st Edition. Sebastopol: O'Reilly Media.

SRIVASTAVA, S. and PRABHAKAR, T. V. 2019. Hospitality of chatbot building platforms. In: *Proceedings of the 2nd ACM SIGSOFT International Workshop on Software Qualities and Their Dependencies.* ACM, pp. 12–19.

TURING, A. 1950. Computing machinery and intelligence. *The Essential Turing: The Ideas that Gave Birth to the Computer Age*, 9: 433–464.

VAIDYA, O. S. and KUMAR, S. 2006. Analytic hierarchy process: An overview of applications. *European Journal of Operational Research*, 169(1): 1–29.

WAGHMARE, C. 2019. *Introducing Azure Bot Service – Building Bots for Business*. 1st Edition. New York: Apress.

WEIZENBAUM, J. 1966. ELIZA a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1): 36–45.

WORSWICK, S. 2019. *Mitsuku chatbot*. [Online]. Available at: http://www.mitsuku.com [Accessed: 2019, August 1].

XU, A., LIU, Z., GUO, Y., SINHA, V. and AKKIRAJU, R. 2017. A new chatbot for customer service on social media. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems.* ACM, pp. 3506–3510.

Contact information

Pavel Kostelník: xkostel2@mendelu.cz
Ivo Pisařovic: xpisarov@mendelu.cz
Mikuláš Muroň: xmuron1@mendelu.cz
František Dařena: frantisek.darena@mendelu.cz
David Procházka: david.prochazka@mendelu.cz