The Botik of Peter the Great:
Creating a Virtual Assistant for the Applicants
to St Petersburg State University

Olga Mitrenina[1[0000-0002-1750-5633]], Maria Ladushina[2[0009-0003-1689-6729]] and Anastasia Kolpashchikova[1[0009-0006-7522-2021]]

[1] St Petersburg State University, 7–9 Universitetskaya Emb., Saint Petersburg 199034, Russia
o.mitrenina@spbu.ru
[2] National University of Science and Technology «MISIS», Leninskiy pr. 4, Moscow 119049, Russia
maria.ladushina@gmail.com

Abstract. This article provides a summary of our research on designing the conversational interface for the Admissions committee of St. Petersburg State University, with a focus on reflecting the university's values and brand. The primary aim was to go beyond simple text-based responses and useful links, and instead, employ a virtual assistant persona embodying Peter the Great to resonate with the modern youth audience and align with the university's historical legacy. To ensure linguistic authenticity, a comprehensive study of Peter the Great's personal correspondence and relevant pedagogical texts from the early 18th century was conducted. Based on these sources, a collection of expressions was curated as a preliminary version of the chatbot's responses. Subsequently, testing was conducted among a group of students to gauge the understandability of the language. Furthermore, a series of studies were carried out to balance the stylistic elements with the requirements of usability. The CAILA platform was used to develop "The Botik of Peter the Great" chatbot, employing language that captures the essence of the Peter the Great era while maintaining an engaging, informative, and user-friendly approach. The chatbot was launched as a widget on the Admissions committee's website and made available on the Telegram messenger platform.

Keywords: Virtual Assistants, Chatbots, Conversational Interface, Language Stylization, Usability Requirements, Middle Russian.

1 Introduction

Dialogue systems are becoming a mandatory attribute of every organization and company that involves interaction with its users. They reduce the workload of customer service employees, as well as make it easier for visitors to find information and reduce the amount of face-to-face interaction. Besides, some users *prefer* using a virtual assistant because they can take their time and ask for information over and over "without feeling like they're 'bothering' a human agent" [1].

Our task was to create a Russian-speaking virtual assistant for the applicants to St Petersburg State University as an addition to the admissions website. The chatbot must act as a widget on the site pages and as an assistant in the Telegram messenger. Its main task is to assist in navigating the site. At the same time, its tasks do not include being a conversation expert, chit-chat bot or similar; its tasks were narrowed down to providing the user with a short textual reply or a useful link.

When developing the bot, it was necessary to combine two contradictory tasks: to be understandable to the modern youth audience, but at the same time to convey the spirit of the oldest Russian university located in the cultural capital of Russia.

2 Selection of assistant's personality

2.1 Why Peter the Great

The first task was to specify a personality for the chatbot.

When developing modern virtual assistants (voice, text, and others), there is a strong tendency to personalize them, giving them traits of a specific personality (chatbot personality, chatbot persona). A personalized virtual assistant begins to fulfill the functions of a company or organization's logo: promoting personalization and distinguishing the marketing object from a myriad of other similar marketing objects, and ideally, visually expressing the main direction of the company's activities and the philosophy of its brand. Unlike an "inanimate" logo, a virtual assistant can interact with visitors, influencing them not only through direct recommendations but also through its speech style.

Saint Petersburg State University is the oldest university in Russia, and Peter the Great is the founder of both the university and the city of Saint Petersburg itself, one of the names of which is "the City of Peter". These facts made it practically inevitable for us to choose Peter the Great as a virtual impersonation of the assistant for aspiring university students. Furthermore, we wanted the chatbot's personality to set a specific tone of communication and prepare students for studying in the oldest university in the country with a rich academic tradition in the "northern capital," renowned for its cultural and historical landmarks, in the most beautiful city of Russia. By making this choice, in construction of our chatbot persona we could use well-known facts about Peter I, primarily him being a great patron of university education, knowledge, and progress in general. The assistant itself quickly acquired the name "The Botik of Peter the Great", given that the word "Botik" in Russian is ambiguous: it means not only the famous boat of Peter the Great, but also a small (chat-)bot.

2.2 Balance Between Past and Future

In the responses of the virtual assistant, we made an effort to adhere to the stylization of the vocabulary and syntax to Peter the Great's era (additional details are discussed in the next section). However, it was necessary to take into account that the bot interacts with a young audience in the 21st century [2]. This determined its "personal characteristics".

- It is polite. More often than not, much more polite than its real-world interlocutors.

- It is "from another era."

- It is aware that it is a robot ("*my technology is still imperfect*"/ "*my achievements are modest, but knowledgeable people tirelessly work on me*" [in order not to overload the text of the article, we provide approximate translations of Russian-language sentences into English]).

- It is scholarly and educated, as evidenced by its rich vocabulary, "bookishness," and a certain eloquence in its speech.

To ensure linguistic authenticity, we consulted authentic texts of Peter the Great's personal correspondence, as well as experts in the history of Peter the Great's era. However, it should be noted that our goal was not to strictly adhere to the language stylization. The very fact that the persona of "Virtual Peter" exists in the modern world and deals with contemporary tasks allowed us to approach the stylization with a certain degree of freedom. In his "backstory," we had to make the assumption that Virtual Peter has familiarized himself with the realities and terminology of the modern era, in order not to impair the user-bot interaction and the success of the user experience. For example, the opening phrase is: *Приветствую Вас, будущий студент! Я Ваш виртуальный помощник с душой Петра I, покровителя наук и всяческого учения. Рад буду Вашей милости ответы на вопросы предоставить.* ('*Greetings to you, future student! I am your virtual assistant with the soul of Peter the Great, patron of sciences and various learning. I shall be honored to guide you on your journey, ask your questions*'). Furthermore, due to the nature of the requested information, it is necessary for the virtual assistant to be familiar with concepts such as telephone, email, website, personal applicant account, and others.

We made an effort to ensure that the content the user seeks is not obscured by archaic speech. Department names, lists of required documents, and similar types of information were conveyed in a modern, neutral language. Considering the specific nature of the information (for example, a user's question might require a large table of specialties or a link to a document with admission conditions), the virtual assistant often responds with a link to a website page, allowing for seamless stylization of the assistant's introductory and concluding statements. The only category we considered potentially affected by such a communication style is foreign students, so in the reply to them, the assistant answers in English first, providing them immediately with the link to the English version of the university website. The message for international students is initially provided in English and without stylization, as we assumed that this group of applicants may experience difficulties in understanding the chatbot.

Subsequently, creating texts for the bot became a separate important task, which was further complicated by the potential conflict between the proposed character and the target audience. Peter the Great is a historical figure, an authoritative personality with established views, a person of his time, from whom we do not expect a mastery of the modern language. In other words, he is a conversational partner who may not be easily accepted by a diverse target audience consisting of applicants. Here, we are not only referring to the diversity in age, knowledge level, and professional skills, but also to the fact that the admission process is often an anxious moment. Therefore, the virtual assistant must not only have a distinct personality but also be comfortable to communicate with, informative, and easily understandable.

3 Language Fine-tuning

3.1 Outline

The language used by the virtual assistant is crucial in representing its unique characteristics and advantages. The linguistic features should always be aligned with the functions of the bot and serve its main purpose.

The process of finding the optimal language solution was conducted in three stages. The first stage involved selecting the main language references and preparing quotations that vividly expressed the peculiarities of the language during Peter the Great's era. The second stage consisted of composing a set of phrases that included stylized responses to address all the key questions outlined in the bot's scenario. During this stage, the initial tests of the virtual assistant's performance were conducted. Based on the results of these tests, the main objective for the third stage became clear – it was necessary to create a balanced language for the chatbot where the stylistic elements would not hinder the fulfillment of its main function. Further details regarding each stage are provided below.

3.2 First Stage: Research

Information was gathered on the language characteristics of Peter the Great's era (early 18th century).

Firstly, an active use of borrowings from foreign languages [3], both lexical and syntactic, was a typical feature of the language during this period. These borrowings were employed to name both new objects and phenomena, as well as familiar ones, much like we currently use terms such as "computers" and "messages."

Another characteristic was the stylistic diversity and even variety, as indicated by numerous researchers [4]. Thanks to these features, we were able to incorporate various elements into the virtual assistant's language, even if they may seem unfamiliar.

Secondly, the frequent use of Slavicisms was prevalent during the early 18th century [4], including words taken from Church Slavonic or constructed according to its model. The stylistic variety was further enriched by their combination with Old Russian words, which were not distinguished by any particular "style" at that time. These words serve as recognizable markers of a different era, and their inclusion became an evident part of our stylization.

As references, we selected Peter the Great's letters [5], as well as the literary and pedagogical monument of the early 18th century, «Honest Mirror of Youth, or a Testimony to Social Intercourse Collected from Various Authors» (second part) [6]. The choice of letters was a reference to the prototype of a virtual assistant, while the second source was taken as an example of Peter the Great's era text, which was created for a wide range of readers, as it contained guidelines for the behavior of "*yeonglings*" and noble girls.

In the letters, the main objective was to identify phrases that could be utilized in the virtual assistant's responses for the following situations:

- Greeting

- Response to a request
- Expression of gratitude
- Response to failure
- Response to the interlocutor's aggression
- Message indicating a lack of understanding of the request

Here are examples of phrases from the letters:

Письмо Ваше приняв, ответствую: ... ('Upon considering your missive I came to the following conclusion…'. Letter 395 to Feodor Apraksin) – For the situation "I understand you, here is the link and an answer to your question."

Письмо Ваше я принял, но не выразумел, засим прошу о том же отписать мне иначе ('Petition of yours rendered me rather confused. Deliver your request in a different manner. Based on a letter 407 to Boris Sheremetiev). – As an example of a classic phrase for a chatbot, "Sorry, I do not understand you".

При сем милости Вашей желаю всякого доброго здравия ('Hereupon I bid you farewell.' Letter 379 to Feodor Apraksin). – Successful assistance or farewell.

The distinctive speech features visible even in the provided examples include archaic words, independent and auxiliary parts of speech (*засим, выразумел, ответствую, зело*), outdated morphemes (*помощию, зачинать*), and word forms (*надеюся*), adverbial participial constructions often found at the beginning of sentences, inverse or unfamiliar word order for modern ears (the definition after the determined word: *письмо Ваше*), the placement of the predicate at the end of the sentence). Additionally, it is noteworthy to mention the spelling of certain words, characteristic of a time when modern orthographic norms were still being established (*естьли*).

Interestingly, during this period, what can now be referred to as the official-business style was beginning to take shape, even evident in the letters of Peter I. Some phrases that may currently appear excessive or humorous could have been considered normal for official correspondence in the first half of the 18th century, and this was also decided to be utilized for stylization.

Unfortunately, suitable examples for responding to the interlocutor's aggression were not found in the letters. This can be attributed to both the genre of the letters and their nature of complementarity, as well as the temporal and spatial distance from the interlocutor.

The reference to the second source, "Youth as an Honest Mirror" [6], provided an understanding that through sentence structure and the incorporation of specific words, one can create the feeling of the language used in the early 18th century without burdening the phrase with obscure terminology.

3.3 A sample of a lexical study

In order to maintain the chosen style, it is necessary not only to add elements that create an impression of 18th-century language but also to remove words and expressions that appear out of place in the context of all the dialogues and which Peter the Great could not have used.

The "backstory" of the bot assumes that the virtual assistant has learned modern words and understands contemporary realities. This is necessary for the chatbot to

remain understandable and useful, as well as able to provide accurate information (about the website, admissions office phone number, document submission method, etc.). However, replacing words that are not directly related to the admissions and education process at the university seems acceptable. Currently, a list of such words is being compiled, and sometimes research into the history of a word is required to determine whether it should be included in this list.

One of such controversial words is the word "*information*." On one hand, it may seem too modern for the 18th century. On the other hand, it has a characteristic Latin appearance, and according to researchers' observations, the number of borrowed Latin and Polish words increased significantly during Peter the Great's era [3].

The 18th-century Russian language dictionary does indeed include the word "*information*" with a note that it is a new word, and with a first occurrence date of 1702, as expected, it indicates its Latin origin and borrowing through Polish (*informacyja*), Ukrainian, and Belarusian (*информация*) languages. However, the definition the dictionary provides differs from what is expected: "*message, notification; report*" (*And as soon as I receive the information, I will send it to Resident Kaster with a special courier.* AB IV 28). It can be assumed that the modern meaning (for example, in "The Great Explanatory Dictionary": 1. *Information*. 2. A report on the state of affairs somewhere, on certain events, processes, etc. 3. Specialized. Information about the state of affairs in the surrounding world, its properties, ongoing processes, etc.) developed later.

Therefore, in the dialogues of the virtual assistant, the word "information" will have to be replaced.

3.4   Second stage: Testing the utterances

Based on Peter the Great's letters, simplified sets of expressions were compiled. For each scenario branch, it was necessary to select several phrases in order to diversify the speech of the virtual assistant. This is required for cases where we anticipate that the user may ask the same question multiple times or begin a conversation with the bot on multiple occasions.

At this stage, the phrases taken from the letters were reworked as follows [7]:

- The syntax of the phrases was simplified, indicating the obvious connection between parts of the sentence and reducing the use of participle constructions. However, the inverted word order for noun + adjective/participle phrases and the placement of the predicate at the end of the sentence were retained.

- The length of the phrases was reduced in order to ensure that the message format does not distract the user from the content.

- Some words have been replaced with more understandable ones. For example: *понял* вместо *выразумел.*

- Religious fragments, which are often found in letters, have been removed: *…предая Вас во всемогущество Божие ('...commending you to the omnipotence of God'), с Божиею помощию ('with God's help')*.

For the remaining parts of the script, utterances in modern language were prepared, after which the word order was changed, outdated words and words with outdated morphemes were added. As a stylistic device, the use of familiar words in atypical positions

was also chosen: ***Такие*** *документы [...]предоставить потрудитесь ('Be so kind as to furnish the following documents')* instead of *эти*.

In addition, it was necessary to come up with response options for obscene language. Users tend to use offensive language towards chatbots more frequently than towards real people, and it was assumed that they would behave the same way even towards a bot in the form of a tsar. In real life, one can end a conversation with an offender or respond sharply, but a bot should maintain the conversation. Therefore, the following formulations were devised as responses to obscene language:

А Вы, я вижу, в древнерусской лексике сильны! Но ежели Вы здесь не за потехой, а за знанием, то извольте вопрошание выбрать ('I gather you have mastered some Russian spells of the days of yore! Yet should you yearn not for amusement but for knowledge, ask your questions with decency').

The intermediate version of the chatbot was tested on a sample of 10 students from the Faculty of Philology and 10 students from the Faculty of Management. After testing, both logs with information about where the user did not want to continue the dialogue and feedback from the students were collected.

The test results showed that in most cases, the chatbot's speech is perceived normally, for example, it is sufficient for it to provide a link. However, some utterances remain unclear, which creates discomfort in communication, even if the necessary information was obtained in the response. There was also a comment about the bot making spelling mistakes in relation to the utterances where outdated spelling of words was used *(естьли)*.

3.5   Stage three: Simplification of the Bot's Language

Considering the test results, the bot's responses were simplified. The utterances were revised to ensure that elements that clearly contributed to the archaic style of speech were retained, but did not cause significant difficulties during testing.

The following stylization methods were preserved:
- Placing the predicate at the end of the sentence;
- Inverting the word order for noun + adjective/participle phrases, although not consistently applied in every sentence;
- Outdated vocabulary mainly retained in the form of function words (e.g. *ежели, сей, покуда*, and a conjunction *да*);
- Outdated vocabulary may be used if the utterance contains no more than two such words and the meaning of the phrase is clear from the context, e.g. *Не для каждой специальности потребно сдавать дополнительное вступительное испытание. Но для некоторых нужно творческое или профессиональное задание выполнить, или же портфолио предоставить.* ('Not every trade requires another trial. Yet some do: be so kind as to provide the full account of your merits and good deeds (portfolio). Otherwise, mayhap, withstanding a creative challenge is in order'*)*;
- The use of familiar words in atypical positions (e.g. *такой документ* instead of *этот документ*);

- Incorporation of idiomatic expressions that are atypical for formal communication to maintain a diverse style, e.g. *Общежития **во всей красе расписаны** на странице сей* ('Our students' abode is vividly presented hither **in all its beauty**.').

4   Technical Implementation

The virtual assistant was created on the Just AI Conversational Platform (JAICP platform) [8], developed by Just AI. It allows the creation of dialogue systems using developing bot scripts and machine learning. JAICP Domain Specific Language (JAICP DSL) was used for some basic requests of the user. It is based on tags and uses the patterns that describe keywords and expressions. For example, here are two of the patterns that attribute the user request to the state "WhatCanYouDo":

```
q!: * {как* вопрос* * [я] мо* [$you] зада*} *
q!: * {$what * $you * (мож*|уме*)} *
```

The main functionality of the bot was implemented using the CAILA (Conversational AI Linguistic Assistant) module [9]. Unlike rule-based approaches that require writing patterns using special syntax, CAILA is based on Natural Language Understanding, including the "self-learning" function on training phrases. This allows, even if just to a certain extent, not to hard-code user phrases but to rely on semantic similarity in user's prompts recognition.

Below is an example of training phrases for the intent for displaying a list of bachelor's programs:

- What are the available majors?
- Find a bachelor's program.
- What exams do I need to take?
- Where can I apply?
- What is the competition for the program?
- How many spots are available for the program?
- I want to study linguistics.
- Choose a suitable program for my studies.
- Where can I read about the programs?
- Where can I get accepted to mathematics and computer science?
- I passed physics and chemistry, where can I apply with them?
- Find a program based on my results in the Unified State Exam (USE).
- I'm applying to the Faculty of Geography, what USE subjects do I need?

The technology works in such a way that the user is redirected to this intent when entering similar (but not exact) phrases. The CAILA evaluates the system's probabilities for different intents.

The most frequent questions were presented in the form of *Частые вопрошания* ('Oft-inquired matters (FAQ)') buttons, which are displayed to the user at the start of the dialogue and after each iteration.

We based our decision on the following advantages of using buttons in text chatbots:
- Buttons significantly reduce the number of failed interactions.
- "User care" as the main principle of UX design: clicking a button is more convenient and/or faster than typing text.
- Buttons are particularly useful when users know what they are interested in but cannot formulate it precisely (e.g., they do not know that all admission campaign events and their dates are referred to as "admission calendar").
- Similarly, buttons help when a dialogue participant does not know what questions other users ask and/or what questions can already be asked, i.e., is not aware of all available information options.

Let us explain the last point: every time a new stage of the admission campaign occurred (the exam schedule was posted, lists of document submitters were published), we added the corresponding question to the top row of the FAQ, thereby simultaneously notifying users about the start of this stage and the availability of information about it.

In the FAQ, we decided not to have a unified syntactic form, formulating questions in a way that a real user might type them into a search engine (some using noun phrases, others using interrogative questions). This was done to mimic the formulations of real users, albeit at the expense of some unity. Later on, we found that frequent user phrases are indeed structured in this way - some queries consist of a phrase as for a search engine, while others are in the form of a question.

Following the logic described in the section on language peculiarities of the bot, we decided not to stylize the language in the FAQ. The final list of "Oft-inquired matters" looked like this:
- Lists of applicants.
- Examination schedule
- Browse undergraduate programs
- Applicant's personal account
- Documents required to apply
- Application dates and deadlines
- Contact the Admission committee
- Browse graduate programs
- Accommodation
- Admission of international students

Since the launch of the bot on June 11, 2023, the developers have constantly analyzed the logs of user conversations and tuned up the system so that it more accurately understands the intentions of the interlocutors. The tuning included correcting tags in JAICP DSL and new trainings for the CAILA using additional training phrases.

5   The Results of the First Admission Campaign

"The Botik of Peter the Great" was available in the summer of 2023 during the University admissions campaign as a widget on the Admissions committee's website [10] and on the Telegram messenger platform @SPBU_Petri_bot.

During the period from June 22 (launch of the chatbot) to September 1, 11,950 unique users communicated with the bot, excluding empty sessions. The total number of messages was 59,573 (excluding empty messages). The average duration of one session was 5 messages per user. The use of the FAQ in form of buttons proved to be justified — in the first two weeks of the bot's operation, 58% of user sessions started with the FAQ button. Phrases from the FAQ were used in 68% of all user prompts, and only 32% of messages were written by users in a free form.

User requests such as "Can you speak English" and "English, please" accounted for 30 messages out of approximately 60,000, which is 0.05%.

Below are the examples of some off topic phrases:

questions like "who are you/what can you do"	575 times
question "how are you doing"	25 times
obscene language	24 times
phrases like "stupid bot"	10 times
phrases of gratitude ("thank you", etc.)	137 times
politely said goodbye to the bot	6 times

Cases of bot responses like "I didn't understand you" account for only about 7% of the total number of messages (3,973 messages out of 59,573). This is a very high response rate.

6   Conclusion

In conclusion, our work on designing a conversational interface for the Admissions Committee of St. Petersburg State University has successfully achieved its objectives. By selecting Peter the Great as the virtual assistant's persona, we were able to not only provide users with concise textual responses and relevant links but also reflect the university's values and brand.

Through an in-depth study of Peter the Great's personal correspondence and pedagogical text from the early 18th century, we ensured linguistic authenticity of the virtual assistant's language. The collection of expressions derived from these historical sources was tested among a group of students to ensure understandability.

Balancing the stylistic elements with usability requirements posed additional challenges, but through careful research and analysis, we achieved a harmonious blend of historical language and user-friendly interaction. The use of the CAILA platform enabled us to develop the "The Botik of Peter the Great" chatbot, which captures the essence of Peter the Great's era while remaining engaging, informative, and accessible.

The successful implementation of the chatbot as a widget on the Admissions Committee's website and availability on the Telegram messenger platform further demonstrates the effectiveness of our language solution in meeting users' needs.

Overall, our research showcases the significance of designing a conversational interface that not only provides practical assistance but also embodies the values and spirit of an institution. The findings from this project contribute to the field of

conversational interfaces and serve as a reference for similar endeavors in utilizing historical personas and language to create engaging and user-centric virtual assistants.

Acknowledgments

References

1. Pearl, C. Designing voice user interfaces: Principles of conversational experiences. "O'Reilly Media, Inc." (2016).
2. Saerbeck, Martin & Schut, Tom & Bartneck, Christoph & Janse, Maddy. Expressive Robots in Education: Varying the Degree of Social Supportive Behavior of a Robotic Tutor. Proceedings of the 28th International Conference on Human Factors in Computing Systems. 3. pp. 1613-1622, (2010).
3. Sanbaiyar V. Inoyazychnaya leksika v putevykh zapiskakh Petrovskoj epokhi [Foreign vocabulary in travel notes of the Petrovsky era. In: Knowledge. Understanding. Skill, no 1, pp. 154-159 (2007)
4. Mayorov A. P. Slavyanizmy v delovom yazyke 1-y poloviny XVIII veka [Slavonicisms in the business language of the 1st half of the 18th century] In: Philological Sciences, no. 4, pp. 29–36 (2009).
5. Bychkov A. F. (ed.) Pisma i bumagi imperatora Petra Velikago [Letters and Papers of Emperor Peter the Great]. (Vols. 2-4). Saint Petersburg: Gosudarstvennaya Tipografiya, (1893-1900).
6. Yunost chestnoe zertzalo ili Pokazanie k zhitejskomu obkhozhdeniyu, sobrannoe ot raznykh avtorov. Napechatasya poveleniem Tzarskogo velichestva [Honest Mirror of Youth, or a Testimony to Social Intercourse Collected from Various Authors. Published by the Decree of His Majesty the Tsar]. Moscow: Khudozhestvennaya Literatura (1976).
7. Babaeva E. A., Zapolskaya N. N. Yazykovoj kontinuum Petrovckoj epokhi: obzor grammaticheskikh traktatov pervoj chetverty XVIII veka [Language Continuum of the Petrovsky Era: An Overview of Grammatical Treatises from the First Quarter of the 18th Century]. Studies in Slavic Historical Linguistics. In Memory of Professor G.L. Khaburgaev. pp. 188-206. Moscow: MSU (1993).
8. JAICP Homepage, https://just-ai.com/platforma-jaicp, last accessed 2023/07/01.
9. CAILA Documentation, https://docs.caila.io/docs/ru/, last accessed 2023/07/01.
10. The website of St. Petersburg State University Admissions committee, https://abiturient.spbu.ru/, last accessed 2023/07/01.