

Т. Новгородский (ГУО «Гимназия № 5 г. Барановичи»)
О. А. Банкевич (научный руководитель)

FEATURES OF LINGUISTIC REPRESENTATION OF AI-GENERATED TEXTS

In recent years, there has been an increasingly noticeable surge in the use of artificial intelligence (AI) technology across various industries. This exponential growth in the popularity of AI can be attributed to its fascinating capabilities, which have proven to revolutionize numerous aspects of our daily lives.

Alongside this the communicative branch of linguistics is actively developing. This branch focuses on the study of language as a means of communication, examining how language is used to convey meaning, establish logical connections and shape utterances. With the rise of AI this communicative perspective has taken on a new significance. Therefore, the particular **relevance** is given to the need for an in-depth study of AI language and text generation techniques.

Taking into consideration the fact that AI is in fact a machine that works according to specific algorithms, it has become quite interesting to comprehensively study AI-generated texts. That's why we have drawn attention to this problem and carried out a comprehensive analysis of the texts generated by AI to develop and improving students' English oral and written speech skills.

In the modern world there is a great variety of chat-bots capable of creating different types of texts. For the research we chose the ChatGPT Telegram bot as the most accessible and popular one.

Since AI is a machine, you need to follow a clear algorithm to write a proper text. As a result of the human-machine collaboration 10 texts were generated. All of them are united by the protagonist – a Belarusian boy named Vasilyok. The protagonist was introduced because he becomes the carrier of the author's ideas and thoughts. After all, it is his perception of the world around him, his feelings and emotions that form the basis of any text. Moreover, students can associate or contrast themselves with the protagonist, which increases interest in studying the text and promotes speech production. Additionally, a prompt was made to formulate titles for those texts.

At first glance, the texts turned out to be quite engaging and informative. Difficulties arose with those texts where knowledge of facts was fundamental. Factual mistakes are a problem with AI-generated texts. Such a text as *The Spirit of Belarus: Tracing Belarusian Traditions (Threads of Heritage)* is extremely disappointing, as only one of the four traditions described in the text is peculiar to the Belarusian people.

The text *Vasilyok and the Magical Land of Belarus: A Tale of Love and Adventure (Discovering the Natural Wonders of Belarus)* was difficult to generate, as to write the text the neural network uses all available on the Internet materials,

neglecting their relevance to reality. So, the text may contain fictitious places, strange geographical objects, etc. Sending repeated prompts with clarifications and requests to correct inaccuracies and deficiencies can take quite a long time.

In the course of the research all 10 texts were analysed in order to understand the quality of the AI generation in terms of the English language.

While studying the theoretical material, the types of analyses accepted in linguistics were scrutinised. But based on the peculiarities of the origin of the texts, taking into account the fact that a deeper, more detailed analysis was required, the following algorithm of analysis, which can be characterised as complex or comprehensive, including the elements of systemic, structural and linguistic analysis, was developed.

The results of the analysis of the AI-generated texts for spelling, punctuation and grammatical mistakes showed that, despite the fact that the language of AI training is English, 10 out of 10 texts contain punctuation mistakes. In this case, the mistakes were related to highlighting circumstantial or adverbial clauses at the beginning of each paragraph.

Searching for some information in open sources on the Internet we discovered that there is a common opinion that AI tries to convey the intonation features of the text using commas to separate secondary parts of sentences from main ones.

Also 10 out of 10 texts contain grammatical mistakes. In one of the texts there is a repeated error in the use of the definite article with geographical names, namely, with the names of the regions of Belarus. Other mistakes are the inappropriate use of the possessive case, the absence of *was/were* when using Past Continuous, the replacement of Past Continuous with Past Simple. In addition, the texts contained unfamiliar expressions the correctness of which had to be studied through various dictionaries. For example, *crispy on the outside, and soft on the inside* or *walking through the streets*.

No spelling mistakes were found in the AI-generated texts.

The data obtained from the analysis can be used both in linguistic and non-linguistic fields of science, such as computer science, cybernetics, neurobiology, psychology and others for studying the functioning, correction of algorithms, training and development of AI.

It should be noted that during the research, a parallel analysis of the AI-generated texts was conducted to ensure their compliance with the requirements set by the English language curriculum for the 6th grade (advanced level). This analysis was carried out because the thematic content of the generated texts corresponds to the topics and the communicative tasks related to the socio-cognitive sphere covered in that year of education.

The research showed that AI-generated texts are far from perfect. However, after adaptation by humans, they can be effectively used for educational purposes to develop or improve oral and written skills. Since these texts are quite patterned (they have a specific structure) that can be fine-tuned by adjusting input

parameters, training data or post-generation editing it is possible to work out a specific algorithm that students can use to structure their monological utterances or write their own stories.

The algorithm for composing an oral or written story was tested on a group of students of the 6th “B” class of the State Educational Institution “Baranovichi Gymnasium № 5”. From January to May 2023 students studied texts generated by the ChatGPT Telegram bot, but adapted by a teacher, in their English lessons.

While working on the texts, pupils’ attention was drawn to the analysis of the text structure, ways of connection of logical parts in the text, the use of conjunctions, adverbial phrases, prepositions, introductory words and constructions, and the variety of linguistic means.

The developed algorithm was used for producing their own monological and written utterances. If at the beginning of the approbation in January the volume and lexical content of oral and written texts in general met the requirements of the English language curriculum for the 6-graders (minimum 9 sentences of 60 words), but logic, consistency and originality left much to be desired, then in May, when writing a text on the same topic again, we noticed the creative approach to the choice of the type and content of the text, significant increase in the volume, use of various connections between parts of texts and language means of expressiveness. Thus, the approbation showed the effectiveness of this algorithm.

Also, during the approbation a survey on the complexity and interest of these texts was held among the 6-graders. It is quite predictable that the level of difficulty and the number of unfamiliar words in these texts almost coincided. The survey also showed that the level of difficulty does not affect the level of interest of these texts.

Additionally, a brochure was developed. That is a set of stories *Inspiring Belarus: Vasilyok’s Tales* generated by AI and adapted by a human. It includes 10 texts with the tasks aimed at developing reading literacy. The brochure can be recommended for use in the 6th grade English lessons in the section of “Cultural Corner” and in elective “Reading Practice” classes.

According to the results of our research, it can be concluded that text generation using neural networks has both advantages and disadvantages. The advantages include

1) a high speed of text generation: the machine generates it in a few seconds, a human will only have to make the draft perfect;

2) creativity and novelty: stories generated by AI can sometimes offer unique and imaginative storylines or plot twists that humans may not have thought of;

3) unlimited content: AI has the ability to generate a significant amount of content, making it a rich source of stories;

4) learning and inspiration: AI-generated texts can serve as learning tools or starting points to stimulate creativity and inspiration;

5) efficiency and accessibility: with AI, you don’t have to search for suitable texts, you can generate them yourself. This saves time and provides instant results.

The disadvantages of the generated texts include the following:

- 1) incorrect writing language: strange grammatical structures, similar sentences;
- 2) monotony of texts (logical, semantic, content and other repetitions);
- 3) unnaturalness and primitiveness of content. Even a huge number of epithets does not save the text, but on the contrary, makes it “clumsy”;
- 4) the text has no depth, does not contribute to the development of critical thinking. AI-generated texts are often used where the topic is simple and unambiguous, and it is not necessary to have some special experience and knowledge;
- 5) presence of serious factual, grammatical and punctuation mistakes in the text;
- 6) necessarily requires human proofreading.

Thus, today AI is one of the most intriguing and rapidly evolving fields of science. Its capabilities in text generation astound the imagination and open up new horizons for various areas of activity. AI-generated text can be a very useful tool in many fields, including education. It can help teachers to create educational materials and tasks of suitable complexity for each student. The automatic creation of teaching aids and various materials can significantly improve the teaching-learning process.

To say that AI will completely replace the human creative potential would be incorrect. However, it can undoubtedly save time and effort for people in creating a high-quality text. Remaining merely a tool, AI is a powerful collaborator that complements and enhances our abilities.

Also it is important to find a balance between human creativity, critical thinking and ethical considerations to ensure that AI serves as a tool that enhances rather than replaces the art of storytelling.