

RU

Разработка учебных материалов по иностранным языкам с применением технологий искусственного интеллекта: возможности и ограничения

Евтушенко Т. Г., Шостак Е. В., Дробчик А. В.

Аннотация. В данной статье рассматривается возможность использования генеративных моделей на основе искусственного интеллекта для разработки учебных материалов. Цель исследования – выявить возможности и ограничения применения инструментов ChatGPT, Perplexity.ai и GigaChat для создания упражнений по английскому языку. Эффективность создания упражнений с помощью генеративных моделей оценивалась экспертами на основе следующих критериев: соответствия сгенерированного моделью результата запросу, простоты создания и необходимости дальнейшего редактирования упражнения преподавателем. Научная новизна данного исследования заключается в оценке эффективности использования ChatGPT, Perplexity и GigaChat для генерации разных типов упражнений, представленных в учебниках авторитетных издательств Oxford и Macmillan (уровней Intermediate и Upper-Intermediate). Анализ результатов показал, что генеративные модели хорошо справляются с генерацией языковых упражнений, направленных на формирование грамматических, лексических и фонетических навыков, составлением тестов с множественным выбором на основе предложенного текста, генерацией идей и коммуникативных ситуаций, а также с созданием текстовых материалов и упражнений на перевод. Применение технологий искусственного интеллекта становится менее эффективным при разработке упражнений, ориентированных на развитие критического мышления обучающихся, и составлении вопросов и утверждений для последующего обсуждения в классе, направленных на развитие умений говорения.

EN

Designing foreign language learning activities using generative artificial intelligence tools: potential and limitations

T. G. Evtushenko, E. V. Shostak, A. V. Drobchik

Abstract. This paper addresses the issue of using generative artificial intelligence (GenAI) tools for the purpose of educational content creation. The aim of the study is to assess the potential and identify the limitations of GenAI in designing English teaching materials. AI-generated content is evaluated based on the following criteria: linguistic-pedagogical accuracy, ease of creation, and need for subsequent revision. The scientific novelty of this study lies in assessing the effectiveness of using AI tools such as ChatGPT, Perplexity, GigaChat for generating diverse exercises for teaching English as a foreign language based on the analysis of authentic tasks from Oxford and Macmillan coursebooks (intermediate and upper-intermediate level). The experimental results suggest that AI tools can be effectively used for generating grammar and vocabulary drills and phonetic exercises, for making glossaries, multiple choice tests, scenarios for communicative situations, as well as creating text materials and translation exercises. However, when it comes to generating freer practice activities aimed at developing students' critical thinking and fostering spontaneous classroom communication, the limitations of using AI tools become markedly apparent.

Introduction

The development of artificial intelligence (AI) had a great impact on educational practices. Advances in natural language processing (NLP) dramatically changed educational experiences, including methods and tools used for teaching foreign languages. The pedagogical scenarios appear to be different nowadays due to the growing possibilities of using automated help while designing curricula, creating tasks, and providing summative and formative assessment.

The relevance of the research is specified by the shift in the educational paradigm from teaching to learning in recent years, and the rapid development of digital tools, in particular, AI-powered tools that enable educators to tailor instruction more effectively and develop engaging activities. Moreover, foreign language teachers in the 21st-century are expected to master information and communication technology (ICT) competencies, i.e. to be able to use ICT effectively in the educational process and understand new educational software, and evaluate its didactic potential (Titova, 2022).

There are a huge number of papers devoted to the use of AI tools in education published on the internet, most of them are not scholarly and consider the problem in general. These are usually real experiences described in tutorials or science popular articles. In spite of the fact that a great number of papers is published on 1) the use of the artificial intelligence and is vividly discussed in a variety of sciences, and 2) the potential usage of LLMs in educational process, and teaching foreign languages, in particular, there is still no precise and detailed description of the AI capabilities in foreign language teaching process, in particular, the accuracy of the output and limitations of designing various classroom activities. These models are just a tool which needs testing and evaluation in the framework of language teaching. Therefore, this study is aimed at testing and analysing the feasibility, real potential and limitations of using LLMs in designing activities to be used in the foreign language classroom.

The research objectives are as follows:

- to estimate the potential of using GenAI for creating exercises in foreign language settings;
- to identify the exercise types for which AI-generated content requires minimal or no revision from an expert (teacher);
- to explore the limitations of using AI tools when designing foreign language classroom activities.

The theoretical framework builds upon existing research in artificial intelligence and language teaching.

The idea of automating different parts of language learning, in particular, the automatic generation of content, is not new. Nowadays when digital tools are integrated into the educational process, it is the instructors' responsibility to ensure that the technology is used to a real purpose (Scrivener, 2011).

Teachers and researchers have been trying to create diverse educational materials using natural language processing (NLP) methods: grammar exercises (Perez-Beltrachini, Gardent, Kruszewski, 2012), vocabulary exercises (Souza, Pereira, 2017), or even earlier multiple choice tests (Mitkov, An Ha, Karamanis, 2006). There is also an attempt to design full courses or curricula and syllabi using AI technologies (Kostikova, Holubnycha, Besarab et al., 2024). The opportunities and convenience and drawbacks of using AI tools in language-based contexts are discussed in detail in (Kasneci, Seßler, Küchemann et al., 2023).

The use of large language models (LLMs) offers some advantages that cannot be ignored in a digital world. Firstly, it promotes a student-centered approach, as it allows to personalize tasks to individual students' needs (Mikolajczyk, Moldovan, Chalupczak et al., 2017). Tasks can be tailored for the specific topic of interest or the specified language level, which gives more precision (Shemshack, Spector, 2020; Bahçeci, Gürol, 2016).

The use of generative models makes it possible to automate or semi-automate the preparation for lessons and the supervision of the educational process, which modifies the procedures in planning educational scenarios (Govindarajan, Christuraj, 2023). Using AI models may be really helpful when used for educational or research purposes in the academic environment (Tuba, Süheyla, 2023).

The weaknesses of using AI models are as such: 1) these materials may need checking and revising/editing afterwards by an expert; and 2) copyright issues arise if the developed items are a part of a commercial project (Poole, 2023).

Taking all this into consideration, there are many ways of improving educational technologies using LLMs. It is still a great challenge to understand all the details of training these models and to be able to detect false information. Nowadays this work is done by a teacher who acts as an expert evaluating the quality of the obtained content, correcting and modifying the output the model provides.

In this work, we use a combination of research methods: qualitative and quantitative analysis of the AI-generated educational content. The potential of models of creating exercises is evaluated using a standardized framework based on task templates found in reputable coursebooks designed for intermediate and upper-intermediate students by authoritative publishers. The reliability is assessed through the expert opinion of two foreign language instructors, with a PhD degree in linguistics and pedagogy. To demonstrate the benefits and limitations of using GenAI tools for designing activities, AI-generated content is assessed according to specific criteria: linguistic-pedagogic accuracy, ease of creation, and time for revision required. Following this a comparative analysis of AI-generated materials and human-created exercises is conducted to evaluate accuracy and identify possible errors, as well as to assess pedagogical quality, creativity and cultural sensitivity of AI-generated educational materials.

The practical significance of the study is determined by the potential to apply the obtained results both in educational settings and for further research. They can be used for designing foreign language teaching materials for coursebooks and online courses, which can lead to the creation of a more personalised learning environment. The integration of AI tools can significantly save teachers' valuable time; however, their effective use depends on understanding the practicality of designing specific tasks with AI tools. It is crucial for educators to discern which tasks should be left for human creativity and which can be efficiently generated by AI tools.

Discussion and results

The existing printed coursebooks for learning activities published by Oxford, Cambridge, Macmillan and others are widely used for teaching languages and have proved their efficiency. However, the challenges of the modern

world motivate us to broaden our horizons and consider new technologies for redesigning the education process to create a more personalised environment.

In our work, we used GenAI chatbots powered by large language models to respond to user's queries. The educational content was generated with ChatGPT, GigaChat and Perplexity. For creating audio and images such online tools as Zvukogram and Shedevrum were used, correspondingly. However, in the current study, more attention is paid to the textual content creation, the latter ones were used to explore the potential of LLMs, not the limitations in creating sound or image, which is left for further research. In the paper, GenAI tools mentioned above will also be referred to as AI assistants, AI tools or generative models.

ChatGPT is one of the most famous LLM. It is based on OpenAI and trained on a vast amount of data (Lund, Wang, 2023). ChatGPT is a service which can easily be accessed through the internet. It can be used for text or code generation in a natural or a formal language. It is also supposed to respond well and fast to the instructions given by a user. The text complexity can be varied in accordance with the needs of the target audience by setting specific parameters in a prompt. Thanks to its fast generation of content and immediate feedback, ChatGPT can be used as a tutor or a partner assisting independent learners and helping with performing language-related tasks (Firat, 2023; Kostka, Toncelli, 2023). It can also be applied for receiving instant written feedback with a list of errors, correction, explanation and suggestions for improving the attached essay (Korenev, 2024).

GigaChat is a chatbot developed by Sber in Russia. Like ChatGPT it was created for generating human-like conversations, giving instant feedback and helping with content creation and information retrieval.

As for Perplexity, it is a free AI-powered search engine (Perplexity.ai), which unlike ChatGPT or GigaChat can provide direct links to sources. While ChatGPT is good at conversations, Perplexity AI also concentrates on study with cited sources.

It is important to mention that we used all three models for generating exercises, however, we did not compare the results across these models.

The research design is as follows.

Step 1. In order to construct a dataset, we start by analysing the structure of units in student's books, New English file, Intermediate level (Oxford University Press, 2018) and Language Hub, Upper-Intermediate level (Macmillan ELT, 2019). At first, we analysed the contents of the coursebooks to identify the essential skills that are to be formed when studying and identified the types of exercises typically provided for language learners, with the sampling of units being done randomly.

Step 2. The activities and instructions being specified and classified we start designing them using AI-powered tools. The experimental input data are obtained by means of creating prompts for generative models such as ChatGPT (version 3.5), Perplexity.ai and Zvukogram.com were used for creating audio materials for the course, while GigaChat and Shedevrum.ai were used to create imaging. Based on the selection, the elements of prompts were developed to enable the AI systems to generate language learning tasks. The generated prompts and queries were introduced into different systems, ChatGPT, Perplexity.ai, etc.; the output generated by the systems was evaluated in terms of effectiveness and relevance of the tasks created.

Step 3. Based on the set of constructed and classified data we made prompts and queries for creating language learning tasks encompassing all the language skills such as speaking, writing, listening, reading, and mediation in order to analyse the potential and pedagogical relevance of creating the exercises for teaching English as a foreign language to students of the intermediate and upper-intermediate level (B1-B2) using generative models.

Overall, the number of prompts using different generative models is 248, iterations and paraphrasing included. The number of key elements (keywords or word combinations) in a prompt such as the target audience, the level according to CEFR, the name of activity itself which is required, the type of content, and the context, etc. varied from 5 to 10. All the generated exercises and text complexity were expected to correspond to B1-B2 level of English knowledge according to CEFR. To validate the correspondence of created materials to the level required, the complexity of all text materials generated by models was measured in an online service text analyzer on a website Roadtogrammar.com.

The activities used in English as a Foreign Language (EFL) coursebooks

As it had been already mentioned, at the first stage of our research we analysed the content provided in Oxford and Macmillan student's books for teaching English as a foreign language to students of an intermediate and upper-intermediate level, both demonstrating a communicative approach to teaching a language. In New English File and Language Hub all the units are divided into 5 main parts such as reading, speaking, listening, grammar, vocabulary, and pronunciation. In each part, texts, exercises, imagery and audio materials are used in accordance with the pedagogical goals and the planned outcome. In all types of activities, the communication is meaningful, i.e. the situations chosen are always real life based, no matter which skills are being trained.

These exercises were classified according to the skills that are being formed and the learning content type proposed to learners, which is shown in the first two columns of Table 1.

In the Speaking section the following content is used: texts, isolated statements, questions for a lead-in activity and for an advanced use of the newly acquired grammar and vocabulary in a discussion, imagery and matching exercises. For teaching reading, the authors used texts, questions, incomplete isolated statements, and matching exercises. Grammar and vocabulary skills are trained and assessed using fill-in-the-gaps, matching, and multiple choice activities. Pronunciation section consists of the images, tables of transcription and matching exercises on the opposition of two sounds. Audio materials and the exercises for pre- and post-listening stages are given in section 'Listening'. In our experimental study, in addition to basic communication modes, translation is also considered, although not presented in the analysed coursebooks. In universities in Russia this form of mediation is used both as a form of formative and summative assessment.

Table 1. *The effectiveness of using AI tools for task elaboration in EFL settings: Grammar, vocabulary and pronunciation*

Skills	Content	Iterations	Accuracy	Need for revision
Grammar	Drills	1	5	no
	Exercises	1	5	no
	Multiple choice test	1	5	no
Vocabulary	Fill-in-the-gap exercises	1	5	no
	Matching exercises (definitions)	> 2	3	yes
	Quiz	> 2	3	yes
Pronunciation	Transcription	1	5	no
	Matching exercises	1	5	no

The performance of GenAI tools in designing EFL classroom activities

Based on a system of tasks offered in the analysed student's books for B1-B2 learners, prompts and queries were generated using AI technologies. The learning content was generated; however, the quality of the output and the accuracy of the response appeared to be different.

To estimate the effectiveness of generating the learning content using generative models, a 5-point scale is proposed. The score '5' stands for 'excellent', which implies that AI models can be easily used for a certain kind of task generation and no polishing afterwards is required. Good '4' implies that the generated output can be easily used by a teacher of English for classroom and self-study activities but needs some polishing. The label '3' stands for 'satisfactory', which suggests that the generated output differs from what was expected, it needs further rephrasing or adding precision in prompts and polishing of the output. '2' implies that the output generated is insufficient, i.e. the results are far behind expectations, it needs rephrasing and thorough correction; more specifications in prompts are necessary, although not exactly clear what kind of precision it requires. '1' stands for those cases which are a complete failure either because of the wrong input given to a system as a prompt or query, or because this task cannot be yet successfully generated by AI systems due to its complexity and still requires human creativity.

As experimental data suggest (Table 1 and Table 2), the accuracy of the output does not exactly correlate with the formation of certain language skills. Nevertheless, it can be clearly seen that the output generated by the model demonstrates a stable good or high quality when designing grammar, vocabulary and phonetic drills (Table 1). As for generating tasks for skills development such as speaking, reading and listening, the results are not that simple and need extended clarification (Table 2).

Table 2. *The effectiveness of using AI tools for designing EFL activities for skills development: Speaking, reading, listening and mediation*

Skills	Content	Iterations	Accuracy	Need for revision
Speaking	Questions	> 2	3/4	yes
	Matching exercise (Warm up)	1	5	no
	Statements	> 2	3	yes
	Images	> 2	2	no
Reading	Text	1	5	no
	Questions for the texts	> 2	2	yes
	Incomplete statements	> 2	2	yes
	Matching exercises (definitions)	5	3	yes
Listening	Audio text (monologue)	1	4	no
	Audio text (dialogue)	1	3	yes
	Audio record	1	5	no
Translation	Sentences in L1 (Grammar)	1	4	yes
	Sentences in L1 (Vocabulary)	1	4	no

The content generated for developing listening skills (Table 2) demonstrates a good quality and accuracy of the obtained materials. Monologue speech is easily generated in ChatGPT or GigaChat and the limitation of the lack of pronunciation practice (Al-khresheh, 2024) can be overcome through other AI resources. The voice is properly recorded in Zvukogram with the right pronunciation, although occasionally with some voices several iterations are required due to the imperfections of the software – it may have some words missed out or intonation might need to be corrected manually. As for generating dialogues and interviews, it is still possible, since Zvukogram offers this option and one can assign different voices including male and female for different speech extracts. The output here seems to be worse, since it implies manual labor on the part of the teacher when formatting the text to be voiced. Prior to Zvukogram one needs to create the dialogue with the help of ChatGPT, GigaChat or Perplexity.ai. Many iterations are used to rephrase the prompt, since the output either is too formal or too informal and does not contain the required grammar or vocabulary which is to be trained through the exercise. In some languages where the register is essential (for example, Russian or Spanish) Perplexity.ai may generate a job interview where the interviewee uses formal address (pronoun usted) and the interviewer – informal address (pronoun tú), which needs further revision on the part of a teacher.

Most problems arise when creating content for developing reading and speaking skills. At this stage of the research, we could not establish the real dependencies between the quality of the output and the formation of reception or production skills.

Table 2 shows that models easily generate matching exercises for lead-in activities providing the instruction and further activities as a bonus which has not been specified in a prompt. However, the generation of prompts for creating various questions that are supposed to elicit the feedback from the target audience or to boost the communicative collaboration (individual, pair or class work) seems to be challenging. On the one hand, a great number of iterations are required, on the other hand, even when a prompt is rephrased and more elements are added, the accuracy of the output leaves much to be desired. One of the difficulties is that the content for the activity is 'overloaded', either the questions or statements are too complicated to understand, or they are too difficult and not interesting to be discussed in groups, which does not encourage a follow-up interaction.

Table 3 shows the output generated by a ChatGPT model in response to a prompt of creating 5 statements for casual talk about men's and women's preferences. It is a good example of the low accuracy of the output. Text complexity of the output makes this task hard to understand instead of motivating them to be engaged in discussions or debates.

Table 3. Prompts and the output for developing speaking skills

Prompt	Output
Write 5 statements about men's and women's preferences in eating which students can discuss	1. Men tend to gravitate towards savory and hearty dishes like steaks and burgers, while women often prefer lighter options such as salads and smoothies. How do you think societal influences play a role in shaping these gender-based food preferences? 2. Research suggests that women are more likely to choose colorful and visually appealing food presentations, while men may focus more on the taste and portion size. How do these differences in food aesthetics impact the overall dining experience for individuals? ---

As for creating prompts for developing reading skills, the easiest task was the generation of texts. As a bonus the basic vocabulary can be highlighted and afterwards given in a list at the end of the output (Table 4). Table 4 shows an example of the output of high quality which does not require two or more iterations.

Table 4. Effective use of GenAI tools for developing reading skills

Prompt	Output
Make up 2 short texts about food habits of 2 different people for EFL students at the B1-B2 level with the highlighted words of the basic food vocabulary.	Meet Sarah, a young professional from the United States. Sarah is a vegetarian who enjoys a balanced diet rich in fruits, vegetables, and whole grains. For breakfast, she likes to have a bowl of oatmeal topped with fresh berries and a cup of green tea. During the day, she often snacks on nuts and yogurt. For lunch, Sarah usually prepares a colorful salad with avocado, cucumbers, and cherry tomatoes. As a sweet treat, she indulges in a piece of dark chocolate in the afternoon. Dinner for Sarah typically includes quinoa with stir-fried tofu and a side of roasted sweet potatoes. Sarah enjoys experimenting with plant-based recipes and loves cooking homemade meals. Key Food Vocabulary: vegetarian, fruits, vegetables, whole grains, oatmeal, berries, green tea, nuts, yogurt, salad, avocado, cucumbers, cherry tomatoes, chocolate, quinoa, stir-fried, tofu, sweet potatoes, plant-based, homemade --- Meet Carlos...

However, it should be noted that the text complexity measured in Text analyzer appeared to be of the beginning of C1 level. The difficulties tend to arise in making exercises for working with the text materials. The questions intended for testing the reading comprehension seem to be created mostly for checking if learners are able to spot and understand certain words from the context, the idea itself is not discussed. Also, these questions or statements are so text-oriented that they cannot be used for further activities eliciting feedback from students and resulting in debates (Table 5).

Table 5. Ineffective use of GenAI tools for developing reading skills

Prompt	Output
We are preparing classroom activities for EFL students at the B1-B2 level. Make up 5 questions which will suit both texts. Make up a matching exercise.	5 Questions for Classroom Activities: 1. What is one dietary preference mentioned in the texts? 2. Name a breakfast item mentioned in both texts. 3. Which country's cuisine is highlighted in one of the texts? 4. What type of dish is paella, and what ingredients are typically used in it? 5. What is a common snack or dessert item mentioned in both texts? ### Matching Exercise: Match the food-related term with its corresponding description. 1. Omelette A. A type of traditional Spanish rice dish 2. Quinoa B. A common breakfast option made with eggs and vegetables

Generating exercises for translation is an easy task for a model, and thus the output can always be used in a teaching process with a little of revision or polishing required from a teacher (expert). The problems arise at an accuracy level,

but they are mostly of pragmatic nature. These mistakes in L1 can be explained by a large number of corpora being in English, therefore the sentences proposed for translation may sound slightly unnatural, as they are a true copy of an English phrase. For example, the passive voice can be wrongly used following the English pattern. However, these errors are easy to correct, and knowing this fact when providing prompts, it is necessary to take into account that 20 per-cent of sentences should be either removed or revised.

The last thing to be considered here is prompting or, to be more precise in-context learning. The model can be trained by the examples/demonstrations we offer it. Working with Perplexity showed good results when we used demonstrations as a part of a prompt. Table 6 shows an example of an exercise for training grammar skills.

Table 6. An exercise for training grammar skills

Prompt	Output
Design two exercises for practicing verbs related to obligation, prohibition, and permission, suitable for B2 students. Follow the pattern. Grammar Complete the extracts with the words in the box. <i>allowed needed obliged ought required supposed</i> 1. I'm not sure. In my school, students weren't _____ to attend lessons. ... 6. You quickly learned what you _____ to do	Exercise 1: Complete the Sentences Complete the extracts with the words in the box. must, allowed, forbidden, have to, permitted, supposed 1. You _____ wear a uniform at my old school. 2. Students aren't _____ to use their phones during class. 3. We _____ finish our homework before coming to school. 4. I was _____ to leave early if I finished my work. 5. Eating in the library is strictly _____. 6. You're _____ to ask questions if you don't understand.

It still requires checking and revision by the expert, nevertheless, it can be revised fast or some of the elements can be removed.

As for developing speaking skills, when asking models for generating similar tasks, the output is doubtful. On the one hand, it can really be used in the classroom, on the other hand, there is not enough creativity from the model. It leaves the idea untouched and changes solely the subject to be discussed (Table 7).

Table 7. An exercise for developing speaking skills (freer practice)

Prompt	Output
Design two more exercises for topic "business" and "media" for B2 students. Follow the pattern. A. PREPARE Work in pairs. Think about all of your previous schools. Discuss rules that you didn't like and why. B. PLAN Work in pairs. Make a list of five rules for the perfect school, for both students and teachers. Use your ideas from Exercise A to help you. C. PRESENT Present your rules to the class. Explain your reasons for each. <i>Our first rule is simple — students must be given a say in the topics and subjects they cover. We think that ...</i> D. DISCUSS As a class, discuss how to improve your language school. Would any of the rules from Exercise C improve your school experience?	A. PREPARE Work in pairs. Think about rules or policies you have encountered in workplaces or companies. Discuss any rules you found unfair or difficult to follow and why. B. PLAN Work in pairs. Make a list of five rules or policies for the ideal workplace, for both employees and employers. Use your ideas from Exercise A to help you. C. PRESENT Present your rules to the class. Explain your reasons for each. Our first rule is that employees should be allowed flexible working hours. We believe this helps improve productivity and work-life balance because... D. DISCUSS As a class, discuss how to improve your ideal workplace. Would any of the rules from Exercise C make a real workplace better? Why or why not?

Thus, generating grammar, vocabulary and phonetic drills and exercises appear not to be a challenging task for AI models. The output that is produced is of high quality and can be used in the classroom without supplementary revision. Besides, the AI tools provide a bonus which might not be necessarily given as a prompt. For example, in addition to the exercises the models also generate the instruction itself, which happens to be rather convenient for a teacher, and a suggestion of further activities for developing productive skills for implementing knowledge and skills acquired in the unit.

As for the activities generated for developing listening, reading and speaking and translation skills, generative models are not so good at creating the learning content corresponding to teachers' needs. Text materials for the required level are easily generated, even with highlighted key vocabulary and a glossary afterwards. The problems may arise when interaction activities for boosting productive abilities are to be created.

There arise no problems in creating testing and translation exercises for formative or summative assessment; however, it is essential to be ready to use only 80 % of the output. The one fifth of generated sentences or questions in L1 might be pragmatically wrong which should be always kept in mind when using a translingual approach.

It is necessary to note that these are not receptive or productive mechanisms which seem to influence the quality of the output. The preliminary results suggest that the factors on which the accuracy of the output generated by AI models in response to a prompt or query, along with completeness and simplicity in designing the task, depends, are specified by the pedagogical goals. In other words, those exercises which require the use of students' cognitive abilities and stimulate critical thinking prove to be the hardest to generate for a model, which can be compensated by changing the number and set of parameters that are included in a prompt. This requires a great number of iterations from a developer (teacher), which might be time-consuming and does not necessarily lead to the accurate and expected output, i.e. the output learning materials still need polishing and revision from an expert.

Based on the obtained results the following recommendations might be given to those teachers who are willing to integrate AI tools into their work. Teachers can use AI models to generate exercises for controlled and semi-controlled practice such as grammar, vocabulary, and phonetic drills as the output is of high quality and requires minimal revision. However, teachers should not forget to review the AI-generated content, especially when it comes to pragmatic accuracy in L1 sentences or questions. At first sight it might look well but deep inside the errors are evident, and students can see it in the classroom. It is up to educators to choose whether they want to remove such items or replace them. Generating texts for reading or listening, scenarios for communicative situations, role plays and instructions for exercises for productive skills to tailor classes more to students' needs seems to be a good idea when the parameters in prompts are chosen properly or the demonstration is provided to the model. Nevertheless, one should be careful with designing activities requiring the use of critical thinking with AI tools. Time spent on adjusting the prompts and multiple model iterations and expert revision is of great value anyway and can be spent for a real purpose.

Conclusion

The results of this research demonstrate that AI assistants are quick-fire tools and the potential of LLMs for creating all types of second language activities is immense. As GenAI tools provide instant feedback, tasks and exercises can be created fast using the required grammar or specific vocabulary in a few seconds. All types of activities for language and skills development used in the analysed Oxford and Macmillan coursebooks are easily generated by ChatGPT, GigaChat and Perplexity.ai. However, the quality of the output does not always meet instructors' expectations. The effective usage of AI tools for designing EFL depends on the teacher's ability to evaluate the pedagogical value of the output, i.e. linguistic-pedagogic accuracy, ease of creation, and time for revision required.

All AI-generated exercises were roughly divided into three groups based on their practicality for educational use: (1) exercises that require checking but no further revision by an expert, (2) exercises needing query adjustment and revision, and (3) exercises requiring multiple iterations and significant expert revision.

The *first group* includes language exercises aimed at forming and developing vocabulary, grammar and phonetic skills. The models also easily generate glossaries, lists of discussion topics and communicative situations with detailed role descriptions within the context of a specific subject.

The *second group* includes the creation of text files aimed at developing reading, listening, and writing skills. AI-generated texts require careful review by an expert to ensure the texts illustrate specific grammatical phenomena. Depending on the exercise's purpose and the model's output, further refinement of the parameters may be needed. This group also encompasses translation tasks for vocabulary and grammatical practice, where in some cases, approximately 20% of the generated sentences may require editing or removal due to inaccuracies in the Russian formulations. Multiple iterations and adjustments are often necessary when producing audio materials. The model handles monologue generation quite well, though iterations might be needed to address missing words or intonational issues. Creating dialogues with different voices assigned to various segments demands more effort and time from the instructor to fine-tune the request parameters.

The *third group* includes exercises that do not facilitate the teacher's workload during lesson preparation. These are exercises for freer practice, namely for developing comprehension or for practicing and extending productive skills. Based on previously generated textual materials, the models generate questions that are often too simple and do not require critical thinking on the reading material. As a result, such questions are ineffective as a basis for discussions in frontal or group work. The accuracy of the outputs significantly declines when generating tasks that require students to synthesise and analyse information. It is also important to mention here that using the demonstration prompt makes the output for all groups more accurate and pedagogically valuable.

The limitations of using AI tools for instructional materials become particularly evident when designing the second and third groups of exercises. AI generated exercises always require careful expert review to ensure accuracy and pedagogical value. Moreover, they often demand significant time investment for prompt and output revision, and they are not always effective for testing comprehension or extending productive skills.

To sum it up, GenAI tools, when used appropriately, can significantly contribute to making learning more engaging and effective. However, due to the complexity and continual improvement of the accuracy of generative models, much remains to be explored. Therefore, a key task for further research is to study the challenges teachers of different languages and language aspects may encounter when designing the educational content. It is of interest to investigate how the generated content varies across different linguistic and cultural contexts. For example, designing educational content in Spanish or French also requires considering the grammatical gender, whereas Russian and German present challenges related to case usage. Thus, it is of importance to study how language-specific features or other factors influence the quality of the generated content.

Источники | References

1. Al-khresheh M. H. Bridging technology and pedagogy from a global lens: Teachers' perspectives on integrating ChatGPT in English language teaching // Computers and Education: Artificial Intelligence. 2024. Vol. 6. <https://doi.org/10.1016/j.caeai.2024.100218>
2. Bahçeci F., Gürol M. The effect of individualized instruction system on the academic achievement scores of students // Education Research International. 2016. <https://doi.org/10.1155/2016/7392125>
3. Firat M. How Chat GPT Can Transform Autodidactic Experiences and Open Education? // OSF Preprints. 2023. <https://doi.org/10.31219/osf.io/9ge8m>
4. Govindarajan R., Christuraj G. Opportunities and challenges of using ChatGPT in the ELT scenario of Utas, Nizwa, Oman // The Journal of Teaching English for Specific and Academic Purposes. 2023. Vol. 11. Iss. 3. <https://doi.org/10.22190/JTESAP230529046G>
5. Korenev A. A. Strategies of using artificial intelligence for written corrective feedback in language education // Lomonosov Linguistics and Intercultural Communication Journal. 2024. Vol. 27. №. 2.
6. Kasneci E., Seßler K., Küchemann S., Bannert M., Dementieva D., Fischer F., Gasser U., Groh G., Günnemann S., Hüllermeier E., et al. ChatGPT for good? On opportunities and challenges of large language models for education // OSF Preprints. 2023. <https://doi.org/10.35542/osf.io/5er8f>
7. Kostikova I., Holubnycha L., Besarab T., Moshynska O., Moroz T., Shamaieva I. Chat GPT for Professional English Course Development // International Journal of Interactive Mobile Technologies. 2024. Vol.18. Iss. 2 <https://doi.org/10.3991/ijim.v18i02.46623>
8. Kostka I., Toncelli R. Exploring applications of ChatGPT to English language teaching: Opportunities, challenges, and recommendations // TESL-EJ. 2023. Vol. 27. Iss. 3. <https://doi.org/10.55593/ej.27107int>
9. Lund B. D., Wang T. Chatting about ChatGPT: how may AI and GPT impact academia and libraries? // Library Hi Tech News. 2023. Vol. 40. Iss. 3. <https://doi.org/10.1108/LHTN-01-2023-0009>
10. Mikolajczyk T., Moldovan L., Chalupczak A., Moldovan F. Computer Aided Learning Process // Procedia Engineering. 2017. Vol. 181. <https://doi.org/10.1016/j.proeng.2017.02.503>
11. Mitkov R., An Ha L., Karamanis N. A computer-aided environment for generating multiple-choice test items // Natural Language Engineering. 2006. Vol. 12. Iss. 2. <https://doi.org/10.1017/S1351324906004177>
12. Perez-Beltrachini L., Gardent C., Kruszewski G. Generating grammar exercises // The 7th Workshop on Innovative Use of NLP for Building Educational Applications, NAACL-HLT Workshop. 2012.
13. Poole F. J. Using ChatGPT to Design Language Material and Exercises // OSF Preprints. 2023. <https://doi.org/10.31219/osf.io/cqnbq>
14. Scrivener J. Learning Teaching. The Essential Guide to English Language Teaching. Macmillan Education, Oxford, 2011.
15. Shemshack A., Spector J. M. A systematic literature review of personalized learning terms // Smart Learning Environments. 2020. Vol. 7. [https://doi.org/10.1186/s40561-020-00140-9\[1\]](https://doi.org/10.1186/s40561-020-00140-9[1])
16. Souza W. I., Pereira Á. R. Jr. A Methodology for Automatic Generation of Vocabulary Exercises // Proceedings of the 23rd Brazilian Symposium on Multimedia and the Web (WebMedia '17)., ACM, New York, NY, USA, 2017. <https://doi.org/10.1145/3126858.3131566>
17. Titova S. V. The Map of Competencies of a Foreign Language University Teacher in the Context of Digitalization of Education // Vysshee obrazovanie v Rossii = Higher Education in Russia. 2022. Vol. 31. № 5. doi.org/10.31992/0869-3617-2022-31-5-133-149
18. Tuba L., Süheyla A. The impact of Artificial Intelligence in academia: Views of Turkish academics on ChatGPT // Heliyon. 2023. Vol. 9. Iss. 9. <https://doi.org/10.1016/j.heliyon.2023.e19688>

Информация об авторах | Author information



Евтушенко Татьяна Геннадьевна¹, к. филол. н., доцент

Шостак Екатерина Владимировна², к. пед. н.

Дробчик Анжелика Витальевна³

^{1, 2, 3} Санкт-Петербургский политехнический университет Петра Великого



Tatiana Gennadievna Evtushenko¹, PhD, Associate Professor

Ekaterina Vladimirovna Shostak², PhD

Anzhelika Vitalievna Drobchik³

^{1, 2, 3} Peter the Great St. Petersburg Polytechnic University

¹ evtushenko_tg@spbstu.ru, ² shostak_ev@spbstu.ru, ³

Информация о статье | About this article

Дата поступления рукописи (received): 06.05.2025; опубликовано online (published online): 06.06.2025.

Ключевые слова (keywords): преподавание английского языка как иностранного; технологии искусственного интеллекта; дидактические материалы для изучения иностранного языка; разработка учебных материалов; большие языковые модели; English language teaching; generative artificial intelligence; language learning materials; content creation in education; large language models.