Liubimov Iurii Vitalevich
student
Olenin Danila Olegovich
student
Shenogin Ilia Alekseevich
student
Plakhotniuk Daniil Sergeevich
student
Scientific adviser
Ioffe Nina Evgenyevna
senior lecturer
MIREA – Russian Technological University
Moscow

USING AI CHATBOTS AS SPEAKING PARTNERS IN TEACHING ENGLISH TO IT STUDENTS

Abstract: *AI chatbots are increasingly offered to language learners as conversation partners, but their quality as tutors is rarely measured. In the study, three large language models available through the Duck.ai service (GPT-6 Luna, Mistral Small 4 and Gemma 4 31B) took part in nine identical role-play dialogues on IT topics in which a B1 learner made 12 typical grammar mistakes. The models corrected 11–12 of the 12 mistakes, but explained the rule in 9–12 cases, and their replies differed in length by almost a factor of two.*

Keywords: *chatbots, artificial intelligence, large language models, English for IT, speaking practice, corrective feedback, language learning.*

Любимов Юрий Витальевич
бакалавр, студент
Оленин Данила Олегович
бакалавр, студент
Шеногин Илья Алексеевич

Content is licensed under the Creative Commons Attribution 4.0 license (CC-BY 4.0)

бакалавр, студент
Плахотнюк Даниил Сергеевич
бакалавр, студент
Научный руководитель
Иоффе Нина Евгеньевна
старший преподаватель
ФГБОУ ВО «МИРЭА – Российский технологический университет»
г. Москва

ИСПОЛЬЗОВАНИЕ ЧАТ-БОТОВ С ИСКУССТВЕННЫМ ИНТЕЛЛЕКТОМ В КАЧЕСТВЕ СОБЕСЕДНИКОВ ПРИ ОБУЧЕНИИ АНГЛИЙСКОМУ ЯЗЫКУ СТУДЕНТОВ, ИЗУЧАЮЩИХ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

Аннотация: *ИИ-чат-боты всё чаще предлагаются изучающим иностранный язык как собеседники, однако их качество как преподавателей редко измеряется. В исследовании три большие языковые модели, доступные через сервис Duck.ai (GPT-6 Luna, Mistral Small 4 и Gemma 4 31B), участвовали в девяти одинаковых ролевых диалогах на IT-темы, в которых учащийся уровня B1 допускал 12 типичных грамматических ошибок. Модели исправили 11–12 ошибок из 12, но объяснили правило лишь в 9–12 случаях, а длина их ответов различалась почти в два раза.*

Ключевые слова: *чат-боты, искусственный интеллект, большие языковые модели, английский язык для IT, развитие устной речи, исправление ошибок, изучение языка.*

Introduction.

Speaking practice is the most difficult part of a foreign language course at a technical university: groups are large, classroom time is limited, and students rarely have a partner to talk to outside the class. According to the interaction hypothesis, learners acquire a language when they take part in meaningful conversation and receive feed-

Содержимое доступно по лицензии Creative Commons Attribution 4.0 license (CC-BY 4.0)

back on their output [5; 7]. Chatbots were proposed as such a partner long before modern artificial intelligence appeared: L. K. Fryer and R. Carpenter noted that bots are always available and do not make shy students feel anxious [2]. A systematic review by W. Huang, K. F. Hew and L. K. Fryer confirmed these advantages and described chatbots as interlocutors for learners, but also pointed to their technological limitations, the novelty effect and cognitive load [3].

Large language models (LLMs) such as ChatGPT changed the situation: they keep a dialogue on any topic and can correct mistakes when asked [4]. For IT students this is especially attractive, because the same tools are already part of their professional work. However, different models are offered to students as if they were equal, while the quality of their feedback is seldom compared. According to R. Lyster and L. Ranta, the type of corrective feedback matters: teachers most often reformulate a learner's sentence (a recast), but recasts rarely lead learners to repair their mistakes, while feedback that draws attention to the error, such as metalinguistic clues, leads to more learner uptake [6].

The aim of this study is to compare how several freely available LLMs act as English speaking partners for IT students: whether they notice and explain learners' mistakes, how long and complex their replies are, and whether they keep the conversation going.

Materials and methods.

Three models from different developers available in the free Duck.ai service in November 2026 were compared: GPT-6 Luna (OpenAI), Mistral Small 4 (Mistral AI) and Gemma 4 31B (Google). For each model, three new chats were opened, one per topic: a job interview for a junior developer, operating systems, and open-source software. Every chat began with the same instruction: «You are my English speaking partner. I am an IT student, my level is B1. Let's talk about the topic I give you. Ask me one question at a time and correct my mistakes.» The learner's level was set according to the Common European Framework of Reference [1].

After the chatbot's first question, four prepared learner replies were sent in each chat regardless of what the chatbot had asked. Together the 12 replies contained 12

Content is licensed under the Creative Commons Attribution 4.0 license (CC-BY 4.0)

typical mistakes of Russian-speaking learners: missing to after want, since instead of for with a period of time, present perfect with yesterday, good in instead of good at, I am use, more easy, two errors of subject-verb agreement (the operating system manage, it boot), a missing verb to be (because it free), can to change, more better and many informations. Thus each model produced 15 replies, 45 in total.

Each mistake was annotated manually as corrected (the reply contained the correct form) and as explained (the reply also named the rule, e.g. «after can, use the base form of the verb»). The texts of the replies were then processed with a Python script that calculated the mean reply length, mean word length, the share of words of seven letters or more, the number of replies ending with a question, and the frequency of IT vocabulary per 1,000 words (a list of 53 IT terms; the count includes the learner's sentences quoted by the chatbots).

Results.

The results are given in Table 1. GPT-6 Luna and Gemma 4 31B corrected all 12 mistakes, Mistral Small 4 corrected 11. The only missed mistake was the first one: Mistral answered «I want become a junior Python developer» with «That's a great goal!» and moved on to the next question. Explanations were given less consistently than corrections: Gemma explained all 12 mistakes, GPT-6 Luna 10 and Mistral 9. In the unexplained cases the chatbots simply rewrote the sentence. For example, both GPT-6 Luna and Mistral turned «Mobile OS needs less RAM and it boot faster» into «Mobile operating systems need less RAM, and they boot faster» and commented only on the plural, so the learner was not told that boot needed an -s. Gemma, in contrast, stated directly that «boot needs an s to match the singular subject it».

Table 1

Comparison of the chatbots as speaking partners
(9 dialogues, 45 replies) (‰ – per 1,000 words)

Feature	GPT-6 Luna	Mistral Small 4	Gemma 4 31B
Mistakes corrected (of 12)	12	11	12
Mistakes explained (of 12)	10	9	12
Replies ending with a question (of 15)	15	15	15
Mean reply length, words	33.9	52.3	65.6
Reply length range, words	14–48	33–92	43–102

Содержимое доступно по лицензии Creative Commons Attribution 4.0 license (CC-BY 4.0)

Mean word length, letters	4.34	4.60	4.41
Words of 7+ letters, %	19.8	25.0	20.8
IT vocabulary, ‰	127.7	96.8	103.7

All 45 replies ended with a question, so all three models kept the conversation going. The replies differed much more in length and style. GPT-6 Luna wrote short replies (33.9 words on average) consisting of one corrected sentence, one short rule and one question, which is close to natural turn-taking. Gemma's replies were almost twice as long (65.6 words); besides the correction and the rule, they contained a reaction to the content of the learner's answer (e.g. «I agree; the terminal is a very powerful tool for developers»). Mistral used the same template in every reply («Almost! The correct way to say this is…», «Correction:», «Question:») and had the highest share of long words (25.0%). Some of its questions tested knowledge rather than invited conversation, e.g. «What is one advantage of a monolithic kernel compared to a microkernel?», which goes beyond a B1 dialogue.

The models also made a few unnecessary changes. GPT-6 Luna removed the article in «the hardware and the memory», which was correct, and Mistral suggested «my opinions» instead of «my opinion». In the sentence «I give you many informations», all three models corrected the noun, but GPT-6 Luna and Gemma also changed the tense (to gave and am giving), interpreting the learner's intention in different ways.

Discussion.

The results show that modern LLMs are able to act as speaking partners: they notice almost all typical grammar mistakes and always continue the dialogue, which overcomes some technological limitations of earlier chatbots noted in [3]. At the same time, the models differ in the type of feedback they give. In about one case out of six, GPT-6 Luna and Mistral corrected a mistake only by reformulating the sentence, i.e. gave a recast, which, according to [6], rarely leads learners to repair their mistakes. Gemma gave explicit metalinguistic feedback in every case, but its long replies may be difficult for a B1 learner and turn a conversation into a grammar lesson. GPT-6 Luna was the most «conversational» partner, while Mistral behaved more like a test.

For teachers, this means that the choice of a chatbot should depend on the aim of the task. For fluency practice, a model with short replies is better; for accuracy work,

Content is licensed under the Creative Commons Attribution 4.0 license (CC-BY 4.0)

a model that explains rules explicitly is more useful. In both cases the teacher should give students a ready role prompt, ask them to request an explanation when a correction is not clear, and remind them that the chatbot can change correct forms. The study is limited by its material: each dialogue was run once, the learner's replies were scripted, the conversation was written rather than spoken, and the behaviour of the models may change with new versions.

Conclusion.

Freely available AI chatbots can serve as English speaking partners for IT students: in nine scripted dialogues they corrected 35 of 36 mistakes and asked a question in every reply. However, the explanation of mistakes, the length of replies and the style of communication differ considerably between models. Chatbots can complement classroom speaking practice, provided that the teacher chooses the model according to the aim and teaches students to use its feedback critically.

Список литературы

1. Common European Framework of Reference for Languages: Learning, teaching, assessment. Companion volume. – Strasbourg : Council of Europe Publishing, 2020. – 278 p.

2. Fryer, L. Bots as language learning tools / L. Fryer, R. Carpenter // Language Learning & Technology. – 2006. – Vol. 10, No. 3. – P. 8–14.

3. Huang, W. Chatbots for language learning – Are they really useful? A systematic review of chatbot-supported language learning / W. Huang, K. F. Hew, L. K. Fryer // Journal of Computer Assisted Learning. – 2022. – Vol. 38, No. 1. – P. 237–257.

4. Kohnke, L. ChatGPT for language teaching and learning / L. Kohnke, B. L. Moorhouse, D. Zou // RELC Journal. – 2023. – Vol. 54, No. 2. – P. 537–550.

5. Long, M. H. The role of the linguistic environment in second language acquisition / M. H. Long // Handbook of second language acquisition / ed. by W. C. Ritchie, T. K. Bhatia. – San Diego : Academic Press, 1996. – P. 413–468.

Содержимое доступно по лицензии Creative Commons Attribution 4.0 license (CC-BY 4.0)

6. Lyster, R. Corrective feedback and learner uptake: negotiation of form in communicative classrooms / R. Lyster, L. Ranta // Studies in Second Language Acquisition. – 1997. – Vol. 19, No. 1. – P. 37–66.

7. Swain, M. Three functions of output in second language learning / M. Swain // Principle and practice in applied linguistics / ed. by G. Cook, B. Seidlhofer. – Oxford : Oxford University Press, 1995. – P. 125–144.

Content is licensed under the Creative Commons Attribution 4.0 license (CC-BY 4.0)