BG Pulse Journal USA
ISSN:1947-2536 http://www.bgpulseusa.com/index.php/bg
BG PULSE JOURNAL

UDC: 811.411.21:37.02:004.8

IMPROVING THE METHODOLOGY OF TEACHING ARABIC AS A FOREIGN LANGUAGE TO UNIVERSITY STUDENTS THROUGH GENERATIVE ARTIFICIAL INTELLIGENCE TECHNOLOGIES

Abdumutallibova Nurkhayobonu Alimardon qizi

e-mail: abdumutallibovanurxayobonu@gmail.com

Abstract: The article examines the methodologically grounded integration of generative artificial intelligence technologies into the teaching of Arabic as a foreign language in higher education. The author proposes a model of teacher-mediated three-stage AI mediation, in which interaction with a language model comprises a teacher-designed prompt, the student's independent dialogue with the model, and a subsequent verification and reflection stage. The model was piloted during the first semester of the 2025/2026 academic year with 55 second-year students. The results demonstrate significantly higher gains in lexical range, oral fluency and learner autonomy in the experimental group. The difference in grammatical accuracy proved less pronounced, which is explained by the language models' errors in vocalisation and in assigning i'rab. Methodological recommendations are offered on this basis.

Key words: Arabic language, foreign language education, generative artificial intelligence, large language model, prompt literacy, learner autonomy, fusha, i'rab, pedagogical experiment.

Introduction

Arabic is spoken today by more than 400 million people and is one of the six official languages of the United Nations. The expansion of trade, economic, educational and cultural ties between Uzbekistan and the Arab states has markedly increased the demand for specialists who know the language. Arabic is currently taught as a first or second foreign language at Tashkent State University of Oriental Studies, the International Islamic Academy of Uzbekistan, the institutes of foreign languages and a number of regional universities.

Nevertheless, mastering Arabic presents Uzbek students with several objective difficulties. First, diglossia is pronounced in Arabic: in the classroom the student studies the literary language — fusha — yet in real communication encounters colloquial dialects. Second, root-and-pattern morphology, the system of i'rab (case endings) and the omission of vocalisation marks in writing complicate text comprehension. Third, the absence in Uzbek of the article, of emphatic consonants and of the category of grammatical gender gives rise to persistent interference errors.

To these is added a problem of a methodological nature. In a group of 15–20 students, over a two-hour practical class each student is left with an average of less than three minutes of oral practice, while the teacher cannot manage a detailed, commented analysis of every written assignment. In other words, the teaching process suffers from a shortage of individualised input and personalised feedback.

The development of generative artificial intelligence (AI) technologies makes it possible to compensate for precisely this shortage. Alongside general-purpose models such as ChatGPT, Gemini and Claude, recent years have seen the appearance of language models specialised in Arabic, such as Jais and Fanar. The "Strategy for the Development of Artificial Intelligence Technologies until 2030", approved by Resolution No. PQ-358 of the President of the Republic of Uzbekistan of 14 October 2024, identifies the introduction of such solutions into the education system as one of the priority directions [1].

Observation of practice shows, however, that teachers and students often use chatbots unsystematically, with the aim of "getting a ready answer". This gives rise to two risks: the language model's errors in vocalisation and in identifying i'rab may be accepted uncritically by the student, and independent thinking may be displaced by excessive reliance on the technology. The problem, then, lies not in the availability of the technology but in embedding it in the learning process on a methodologically grounded basis.

The aim of the research is to develop a methodological model for the use of generative AI technologies in teaching Arabic as a foreign language and to substantiate its effectiveness through a pedagogical experiment.

Literature review

Research on the role of generative AI in language education has grown rapidly since 2023. The work of L. Kohnke, B. Moorhouse and D. Zou summarises the principal affordances of ChatGPT for language learning as the generation of unlimited input, the provision of personalised feedback and the performance of the role of a conversation partner; at the same time the authors stress that these affordances will not be realised unless the digital competence of both teacher and student is sufficient [2, p. 540].

D. Zou and co-authors carried out a systematic review of 49 empirical studies published in 2023–2024 and established that generative AI performs five distinct roles in the classroom setting: source of feedback, tutor, cognitive stimulator, mediator of interaction, and interlocutor. The reviewers point out as limitations that the bulk of the studies address English and written production, and that the results frequently rest on students' self-assessment [3].

Among the works conducted directly on Arabic-language material, the study by S. Humairoh, N. Hadi and U. Hanifah deserves attention. They distinguish three pedagogical functions of ChatGPT in Arabic language education — delivering interactive language material, providing corrective feedback on errors, and working on a text as a co-author — and conclude that the use of the model must in every case take place within a clearly defined pedagogical framework [4, p. 15].

Theoretically, this approach draws on L. Vygotsky's concept of the zone of proximal development: the language model serves as temporary scaffolding for tasks that the student cannot perform independently but can accomplish with support. S. Krashen's comprehensible input hypothesis and M. Swain's output hypothesis likewise provide a methodological basis for interpreting interaction with AI, since the chatbot appears simultaneously as a source of text at the i+1 level and as a testing ground for the learner's own output.

In the domestic literature, the integration of artificial intelligence into education has been treated mainly from a general didactic or technological standpoint; studies devoted specifically to the methodology of teaching Oriental languages to an Uzbek-speaking audience are almost non-existent. The research gap is therefore defined as follows: although the general affordances of generative AI in language education have been studied in sufficient depth, no methodological model has been developed that takes into account the specific features of Arabic — diglossia, the system of i‘rab and vocalisation — together with the interference arising from Uzbek.

Research methodology

The research was conducted in the first semester of the 2025/2026 academic year with the participation of second-year students in the Oriental philology programme of a higher education institution. The sample comprised 55 students in total: the experimental group (EG) — 28 students, the control group (CG) — 27 students. On the results of the entry test at the beginning of the semester the groups were found to be statistically homogeneous (proficiency level — the A2–B1 range).

The experiment lasted 16 weeks and covered four hours of practical classes per week. In the control group, classes were conducted traditionally on the basis of the current curriculum. In the experimental group, approximately one third of class time, together with the independent-study assignments, was organised on the basis of the model of “teacher-mediated three-stage AI mediation” developed by the author.

The essence of the model is as follows. At the first stage, proceeding from the aim of the lesson, the teacher designs a model prompt (query) that explicitly specifies the language level, the length of the text, the topic, the register and the constraints. At the second stage the student works independently with the language model — obtains a text, asks questions, holds a dialogue, has his or her own written work analysed. At the third stage the material obtained is checked in class in the presence of the teacher: the student is required to find at least one error made by the model and to justify it by reference to a grammatical rule. It is precisely this third stage that prevents blind reliance on the technology and returns control over the material being studied to the student.

Five types of prompt were systematised in the course of the experiment (Table 1). Each of them is tied to a particular didactic task; at the beginning of the semester the students were given a two-hour instructional session on the composition of prompts.

Table 1

Classification of the prompts used in the AI-mediation model

Prompt type	Didactic function	Sample task
Input	Generating a level-adapted text on a given topic	“Compose a 120-word fusha text on the topic ‘At the market’ for an A2-level student and add the vocalisation marks.”

Prompt type	Didactic function	Sample task
Drill	Consolidating morphological and syntactic skills	"Conjugate the following verb in the past tense across all 14 person–number forms and give an example sentence for each."
Dialogic	Extending oral communicative practice	"You are a shopkeeper in Cairo and I am a customer. Answer only in fusha and list my mistakes at the end."
Corrective	Commented analysis of errors	"Find the i'rab errors in my essay and explain the reason for each one in Uzbek."
Reflective	Self-assessment and planning	"Based on my mistakes this week, draw up a three-day revision plan for me."

The following methods were used in the research: comparative analysis of theoretical sources; pedagogical observation; entry and final testing; semi-structured interviews with students; a questionnaire survey; and statistical processing of the data obtained by means of Student's t-test. Assessment was carried out against four criteria on a 100-point scale: lexical range, grammatical accuracy (primarily in terms of errors of i'rab and verb conjugation), oral fluency, and level of learner autonomy.

Analysis and results

The results of the pilot experiment are presented in Table 2. The indicators are given as group mean scores.

Table 2

Initial and final indicators of the experimental and control groups (mean score)

Criterion	EG initial	EG final	CG initial	CG final
Lexical range	54.3	71.8	53.7	61.2
Grammatical accuracy	51.6	63.2	52.1	59.4
Oral fluency	49.8	68.4	50.3	57.6
Learner autonomy	46.2	70.1	45.9	51.3

Criterion	EG initial	EG final	CG initial	CG final
Overall mean	**50.5**	**68.4**	**50.5**	**57.4**

As the table shows, the overall mean indicator in the experimental group rose from 50.5 to 68.4 points, that is by 17.9 points, whereas in the control group this increase amounted to 6.9 points. The statistical significance of the difference was tested by Student's t-test, yielding $t = 3.84$ (53 degrees of freedom, $p < 0.05$). This means that the gain in the experimental group cannot be regarded as due to chance.

The most substantial shift was recorded on the criterion of learner autonomy (+23.9 points). According to the final questionnaire, 82 per cent of the students in the experimental group reported that by the end of the semester they were regularly performing written or oral practice in Arabic outside class hours; at the beginning of the semester this figure stood at 29 per cent. In the interviews, students particularly emphasised that they were not embarrassed to make mistakes in front of an artificial interlocutor and therefore attempted to speak more. This finding is explained by a lowering of the psychological barrier and also partly accounts for the +18.6-point gain achieved in oral fluency.

The gain in lexical range (+17.5 points) is chiefly connected with the model's capacity to reformulate the same content in different registers: students could obtain three or four variants of a single sentence and transfer the synonymic set into their active vocabulary.

At the same time the difference in grammatical accuracy turned out to be smaller than expected (+11.6 against +7.3 points). Several reasons for this were identified. First, language models make systematic errors in placing vocalisation marks and, especially, in assigning i'rab endings — analysis of the 240 model responses collected during the experiment showed that in 31 per cent of cases at least one vocalisation or i'rab error was present. Second, even when not asked to do so, the model sometimes mixes lexical items characteristic of the Egyptian or Levantine dialects into a fusha text. Third, the model explains an error "with confidence", which misleads a student at elementary level.

It is precisely this state of affairs that substantiates the principal methodological conclusion of the research: in the Arabic language classroom generative AI yields results not as an autonomous teacher but only as a practice tool under the teacher's supervision. When the third — verification — stage was removed (this was tried out after half a semester, for the sake of comparison, in a small subgroup), the students' ability to notice errors did not improve; on the contrary, they displayed a tendency to accept any answer given by the model as the truth.

The observations also showed that the skill of composing a prompt is becoming an independent methodological object. When the student specifies the language level, the length of the text and the register precisely in the query, the instructional value of the material supplied by the model increases appreciably. It is therefore advisable to formalise prompt literacy as a component of the practical course in Arabic.

Conclusions and recommendations

The results of the research conducted allow the following conclusions to be drawn.

First, generative artificial intelligence technologies are an effective means of overcoming the shortage of input and personalised feedback in teaching Arabic as a foreign language. The pilot experiment confirmed that the overall attainment indicator in the experimental group grew more than twofold in comparison with the control group.

Second, what is decisive is not the technology itself but the manner in which it is embedded in the learning process. The three-stage model — a teacher-designed prompt, the student's independent dialogue, and verification and reflection in class — brings the use of the model under control and subordinates it to the instructional aim.

Third, the errors of language models in assigning i'rab and vocalisation marks remain, for the time being, a serious limitation. For this reason, when working on grammatical material the model's answer must without fail be checked against a textbook or a dictionary.

Fourth, the formation of prompt literacy should be included in the curriculum as a separate module of the practical course in Arabic. In it the student learns not only to compose a query but also to evaluate the answer received critically.

On this basis the following recommendations are put forward: to include in the working programme of the practical Arabic course a module of at least four hours on "Digital tools and prompt literacy"; to redesign part of the independent-study assignments on the basis of the AI-mediation model; to document the results of students' work with AI in the form of a portfolio and to recognise it as one of the assessment criteria in interim assessment; and to prepare a methodological handbook for teachers on the use of language models. As a further prospect for the research, the application of the model to the practice of written translation and of work with specialised (religious, legal, medical) texts has been identified.

References

1. Resolution No. PQ-358 of the President of the Republic of Uzbekistan of 14 October 2024 "On approval of the Strategy for the Development of Artificial Intelligence Technologies until 2030" // National Database of Legislation. — Tashkent, 15.10.2024.

2. Kohnke L., Moorhouse B.L., Zou D. ChatGPT for Language Teaching and Learning // RELC Journal. — 2023. — Vol. 54, No. 2. — P. 537–550. DOI: 10.1177/00336882231162868

3. Zou D. et al. Generative AI (GenAI) in the language classroom: A systematic review // Interactive Learning Environments. — 2025. DOI: 10.1080/10494820.2025.2498537

4. Humairoh S.W., Hadi N., Hanifah U. The AI Revolution in Arabic Language Learning: An Analysis of ChatGPT's Role in Autonomous Learning // Alibbaa': Jurnal Pendidikan Bahasa Arab. — 2026. — Vol. 7, No. 1. — P. 1–19. DOI: 10.19105/ajpba.v7i1.22479

5. Vygotsky L.S. Mind in Society: The Development of Higher Psychological Processes. — Cambridge, MA: Harvard University Press, 1978. — 159 p.

6. Krashen S. Principles and Practice in Second Language Acquisition. — Oxford: Pergamon Press, 1982. — 202 p.

7. Swain M. Three functions of output in second language learning // Principle and Practice in Applied Linguistics. — Oxford: Oxford University Press, 1995. — P. 125–144.

8. Godwin-Jones R. Distributed agency in second language learning and teaching through generative AI // Language Learning & Technology. — 2024. — Vol. 28, No. 2. — P. 5–31.

9. Ryding K.C. A Reference Grammar of Modern Standard Arabic. — Cambridge: Cambridge University Press, 2005. — 723 p.

10. Wahba K.M., Taha Z.A., England L. Handbook for Arabic Language Teaching Professionals in the 21st Century. — New York: Routledge, 2018. — 402 p.

11. Alhaisoni E., Alhaysony M. An Investigation of Saudi EFL University Students' Attitudes towards the Use of Mobile Technologies in Language Learning // International Journal of English Linguistics. — 2017. — Vol. 7, No. 1. — P. 91–99.

12. Kukulska-Hulme A., Viberg O. Mobile collaborative language learning: State of the art // British Journal of Educational Technology. — 2018. — Vol. 49, No. 2. — P. 207–218.