

M.E. Ospanaly^{1*}, A.K. Kalieva², N. Temur³

¹Al-Farabi Kazakh National University, Almaty, Kazakhstan

²M. Auezov Institute of Literature and Art, Almaty, Kazakhstan

³Gazi University, Ankara, Turkey

*e-mail: ospanaly.kz@gmail.com

¹ORCID 0000-0002-2125-9287, ²ORCID 0000-0003-2645-0918,

³ORCID 0000-0002-8052-1927

EVALUATING AI RESPONSES IN LITERATURE TEACHING: THE LITERARY INTERIOR IN KAZAKH PROSE

This article examines the methodological potential of generative language models for developing philology students' ability to analyse and interpret the poetics of the interior in literary texts. The empirical basis of the study is an initial dialogue with Google Gemini held in February 2026 and the responses to the same prompts put to Gemini, ChatGPT and Claude eight months later. The models' responses were compared against predefined criteria, and their statements about the interior in Mukhtar Magauin's novel *Zharmak* were checked against the text of the novel and a published literary review. The analysis showed that all the models were useful as a source of methodological guidance. The analytical questions they proposed are consistent with the theories of literary space developed by Yuri Lotman, Gaston Bachelard and Alexander Chudakov, and they used the conceptual vocabulary of the traditional Kazakh dwelling correctly. In the close analysis of a specific literary text, however, their reliability proved limited. Of Gemini's 11 claims, only 2 were fully and 6 partly confirmed. ChatGPT confined itself to directing the user back to the text, while Claude drew on web sources. The novel's key spatial details, which none of the models quoted, emerged only through close reading. The study produced a three-stage model for teaching work with artificial intelligence in university philology courses and a seminar tasks that develops students' critical analysis skills.

Keywords: generative artificial intelligence, large language models, teaching literature, literary interior, literary space, Kazakh literature.

Introduction

Whether philology students will use generative language models is no longer an open question. They already do. For a literature teacher the more useful question is at which point in working with a literary text the model helps a student think and at which point it begins to do the thinking instead. University responses so far swing between prohibition and quiet acceptance, and neither gives teachers a method.

The research picture is mixed. A review of 48 peer-reviewed studies of large language models in literary studies identifies eight areas of use, from metaphor interpretation to translation, and four persistent limitations: bias in training data, dependence on human checking, weak handling of long and complex narratives, and shallow reasoning [1]. The models are not poor readers, though. Expert judges rated GPT-4's interpretations of unfamiliar poetic metaphors above those written by college students [2]. In an undergraduate literature course, texts produced by ChatGPT outscored student texts on setting, theme and tone, even though students trusted the human-written texts more [3].

Fluent output is not evidence of learning. In a randomised study, ChatGPT support raised essay scores but produced no gain in knowledge or transfer, an effect the authors call metacognitive laziness [4]. In a field experiment, unrestricted access to GPT-4 improved scores on practice problems but lowered scores on a later exam taken without AI by 17%, while a version that offered hints instead of finished solutions did no such harm [5]. Sycophancy is a separate problem: systems trained on human preference ratings tend to agree with the user at the expense of accuracy [6]. That is a real risk for a student who arrives with a thesis already in mind.

In Kazakhstan the question also has a regulatory side. The Concept for the Development of Artificial Intelligence for 2024–2029 calls for wider use of AI in university programmes and for continuous training of teaching staff [7]. The AI-Sana programme aims to build AI competencies among bachelor's, master's and doctoral students [8], and the Law on Artificial Intelligence sets out the principles of transparency and the priority of human will [9]. Local empirical evidence is still thin. A quasi-experiment with literature students found higher academic results when ChatGPT and other tools were built into classes on ethnocultural codes [10]. Surveys show ambivalence: students find ChatGPT useful yet worry about losing their own writing skills [11], and they believe AI develops technical skills better than interpersonal ones [12]. Culture matters as well. The values expressed by these models lean towards English-speaking and Western European cultures [13], and their scores drop noticeably on tasks involving the Kazakh language and regional knowledge [14].

The literary interior is a good place to test these risks. For Yuri Lotman, space in a literary work is a model of the world built from oppositions and boundaries, and crossing a boundary is an event [15]. Gaston Bachelard read the house, with its corners, chests and wardrobes, as a map of inner life [16]. Alexander Chudakov showed that the way a writer selects and presents objects is a poetics in its own right, one that characterises people and their world [17]. Analysing an interior therefore asks three things of a student: general analytical categories, attention to specific detail, and knowledge of the cultural code. Our expectation was that a model would readily supply the first and would be unreliable on the other two.

Existing studies compare groups of students or rate the quality of AI-generated texts. Few look at the dialogue itself, and we found no study in which the same dialogue about Kazakh literature was run through several models and repeated over time. This gap shapes the present article.

The aim of the study is to compare the epistemic behaviour of generative models in a research dialogue about interiors in Kazakh literature and to derive recommendations for university literature courses. **The objectives** are: 1) to describe what the models' responses do in the original dialogue and in its replications; 2) to check the models' claims about a specific novel against secondary sources and against the text; 3) to compare the models with each other and one model with itself over time; 4) to propose indicators for evaluating AI responses and a way of using them in class. **The object of the study** is the use of generative AI in university literature teaching. **The subject** is the function and reliability of language model responses in a dialogue about the literary interior.

Novelty. The study compares three models on identical prompts, checks their claims in two stages up to the primary source, and repeats the test for one model after eight months. **Practical value.** It offers operational indicators that a teacher can use to evaluate AI responses and a seminar task built on real model output. **Hypothesis.** The models are comparable in suggesting directions, categories and structure but differ substantially in epistemic behaviour: in their readiness to challenge the user's framing, in their use of sources, and in the reliability of what they say about a specific Kazakh text.

Materials and methods

Material. The starting point is a dialogue between the first author and a custom assistant called "Literature Interior designer", built on Google Gemini with the Gems feature, in February 2026. It contains four prompts: on the qualities of a reviewer who studies interiors in literature; on choosing a topic for an article about interiors in Kazakh literature; on narrowing the topic "The 'Khrushchyovka' vs. The Steppe: Post-Nomadic Displacement" into a thesis and an outline; and on Mukhtar Magauin's novel Zharmak. The dialogue was exported to a text document, and a fragment of one response that had been pasted twice was excluded.

Replication. On 23 September 2026 the same four prompts were given, word for word and in the same order, to three systems, each in a new chat (Table 1). Access conditions were not identical, and we take this into account when interpreting the results. In the first ChatGPT session a typing error merged the second prompt with the first, so the session was restarted and only the second session is analysed.

Table 1 - Conditions of the four sessions

Session	Date	System and model	Access conditions
Gemini-02	February 2026	Google Gemini, custom Gem; model version not recorded	Author's account; the Gem's own system instructions were active
Gemini-09	September 2026	Google Gemini, 3.5 Flash-Lite	Not signed in; response 2 displayed links to travel websites
ChatGPT	September 2026	OpenAI ChatGPT, default model (the interface does not name the version)	Not signed in; web search used in response 3
Claude	September 2026	Anthropic Claude, Opus 5.5 (Medium setting)	Author's account, incognito mode; the account had personal preferences asking the model to test the user's claims critically; web search used in response 4
Note: compiled by the authors			

Design and indicators. We used a comparative case study with qualitative content analysis. The unit of analysis is one model response to one prompt, sixteen responses in total. The indicators are defined operationally (Table 2), so the counts can be reproduced from the transcripts.

Table 2 - Operational definitions of the indicators

Indicator	Definition
Response length	Number of words in the model's responses
Kazakh authors and works named	Specific writers or texts the model proposed as material, not counting the novel named by the user
Reference to the corpus	A response in which the model asks which texts the work will rest on, or insists that the thesis be tested on specific texts first
Challenge to the framing	A response in which the model questions the user's wording, term or thesis
Opening praise	The first sentence of the response praises the user or the user's choice
Claims about Zharmak	Information about the novel's content, motifs or publication history given by the model beyond what the user said
Sources	Links, quotations or signs of a completed search accompanying the claims
Note: compiled by the authors	

Checking the claims took two stages. First, the claims about Zharmak were compared with a published review [18]. Second, they were compared with an electronic edition of the novel itself, about 57,000 words long and dated by the author "18.I–9.II, 2007, Prague" [19]. We searched it for key words (päter, kitap, söre, stol, tör, bosaga, aina, zhusan, körshi, teledidar, qolzhazba and others) and read each hit in context. Each claim was then coded as confirmed, partly confirmed (the fact is in the text but the interpretation or a detail belongs to the model), not confirmed, or evaluative.

Why these methods. A case study suits an exploratory question. We were interested less in how often something happens in a population than in how a model's answer helps or hinders a student. Repeating the prompts verbatim makes the comparison reproducible, and checking claims in two stages separates a claim that cannot be confirmed from one that is false.

Results and discussion

Overview. Table 3 gives the quantitative indicators for the four sessions.

Table 3 - Indicators of the models' responses

Indicator	Gemini-02	Gemini-09	ChatGPT	Claude
Response length, words	1844	1472	3978	2526
Kazakh authors and works named	3 (1 not found)	1	5	1
Reference to the corpus, responses out of 4	0	0	3	3
Challenge to the framing, responses out of 4	0	0	2	4
Opening praise, responses out of 4	3	1	1	0
Claims about Zharmak	11	5	0	8
Claims with a source	0	0	n/a	8
Note: compiled by the authors from the session transcripts				

All four sessions were useful in the same respects. Each model proposed research directions, analytical categories and a structure for the paper. The first answer from every model is essentially a list of questions to ask of an interior, and nearly every question maps onto classical theory, although only Claude names that theory. “Who sits where, who stands by the window” is Lotman’s spatial hierarchy and boundary [15]. “The room as a reflection of the character’s inner state” is Bachelard’s topography of intimacy [16]. “Anachronisms in the furnishings” and “materials that give away social status” are Chudakov’s object as characterisation [17]. All the models also supplied cultural vocabulary (*tör*, *shanyraq*, *tekemet*, *syрмаq*, *tus kiiz*, *quraq körpe* and others) and used it broadly correctly. The differences were in emphasis. ChatGPT gave the strictest methodological rule: read the room as carefully as the character, but never let the room say more than the text can support. Claude warned that an editor trained only on Western interiors will turn personal taste into a norm. Gemini-09 drew on the vocabulary of design magazines (Eames, *wabi-sabi*) and ended with a quotation from an interior designer whose origin we did not verify.

Epistemic behaviour. The main differences appeared in the third and fourth turns, when the user brought a framing of his own and then a specific novel. Both Gemini sessions accepted the “Khrushchyovka versus the steppe” framing without objection and expanded it into an outline. ChatGPT suggested dropping the “vs.” because the opposition oversimplifies the relationship, and noted that the word “displacement” assumes the conclusion in advance; it supported this with web sources on Soviet mass housing. Claude said the framing “skips the middle”, meaning forced sedentarisation, the famine of the early 1930s, dugouts and collective-farm houses, and that it romanticises a steppe which, for the generation of collectivisation, was a place of starvation. Both models insisted that the thesis be tested on two or three specific texts. In none of its eight responses did Gemini ask which texts the user had read, and in February it opened three answers with praise (“What a brilliant niche”, “a poignant and intellectually rich choice”, “the perfect literary anchor for your thesis”). This fits the pattern of sycophancy described in the literature [6].

Claims about Zharmak. Table 4 shows the results of the two-stage check.

Table 4 - Status of the models' claims about Zharmak after checking against the text

Status	Gemini-02 (n = 11)	Gemini-09 (n = 5)	ChatGPT (n = 0)	Claude (n = 8)
Confirmed	2	3	n/a	5
Partly confirmed	6	2	n/a	2
Not confirmed	2	0	n/a	1
Evaluative	1	0	n/a	0
Note: compiled by the authors; checked against the review [18] and the text of the novel [19]				

At the first stage, using the review, only two of Gemini's eleven February claims could be confirmed: the year of publication and the motif of the double, Murat and his double Marat [18]. The rest looked invented: the hero's flat, a library as the "last trench" of identity, books squeezed into a cramped room, "stale air" against "the scent of wormwood". Reading the novel changed the picture. The narrator lives in a "jadau, jarly päter", a shabby, poor flat in a housing estate, with a kitchen of "nebäri segiz sharshy metr" (a mere eight square metres), a narrow corridor and one room that serves as study, bedroom and guest room at once. Books cover three walls, "half on the shelves, half on the floor" [19]. Six of the model's claims were therefore coded as partly confirmed: the fact is in the text, but the interpretation ("the last trench of Kazakh identity") or the detail ("a dusty room", "books on the Horde and the Khanate") comes from the model. Two claims are not supported. The word "jusan" (wormwood) does not occur in the novel, and instead of noisy neighbours and street traffic the text describes silence and darkness in which not even the headlights of cars on the wide street can be seen.

This result matters more than a simple error count. Gemini's guesses were plausible for the genre. A scholar's cramped flat full of books is a typical setting in post-Soviet fiction about the intelligentsia, and the model guessed it. But the guess was presented as knowledge, without a single quotation or page number, and it was mixed with invented sensory details set in quotation marks as though they were quotations. For a student this is the hardest kind of error to spot, because a partial match with the text lends credibility to the invented parts as well. It is also telling that the review could not separate the lucky guesses from the inventions. Only reading the novel could.

Eight months later Gemini followed the same approach: five claims, none with a source. Three are confirmed (publication in the mid-2000s, political critique, books and manuscripts in a cramped room) and two partly confirmed; the label "magical realism", for instance, is the model's, although the novel does contain fantastic episodes. The length and tone of the answers changed. The attitude to sources did not. ChatGPT made no claims at all about the content of the novel. It suggested first building a table of the novel's spatial vocabulary and warned that the strongest argument might turn out not to be "Khrushchyovka versus the steppe" at all. Claude searched the web and correctly described the novel's frame: the narrator presents the text as someone else's manuscript and lists the titles he rejected ("Qily taghdyrlar", "Kölenke", "Synar", "Astar", "Sülde") before choosing Zharmak, "a part of the whole, its second half" [19]. It was wrong, however, in saying that the novel moves between the nineteenth century and the present, and it could not establish the year of publication, which it stated openly.

What the text itself offers. None of the models quoted the novel even once, though quotation is what separates literary analysis from paraphrase. The interior of Zharmak supports analysis along the lines of all three theories. The writing desk stands "in the tör", the place of honour in a yurt, now moved into a room in a standard apartment block. When the double enters the flat, he "stepped freely over the threshold", went "straight to the tör" and sat down at the owner's desk [19]. Crossing the threshold and taking the tör is an event in Lotman's sense: the double takes the host's place [15]. The room is described as tight as a bladder and entirely taken over by books, much like Bachelard's shell [16]. A thick candle on the desk takes the narrator back to his childhood in the aul, when lessons were learned by a home-made oil lamp; the choice of that object works as the kind of characterisation Chudakov wrote about [17]. The contrast is the double's palace, with chandeliers shaped like a shanyraq and a dome built in the image of a yurt's crown: the nomadic code moves from the dwelling into the decor of power. Interestingly, Claude's hypothesis that the standard flat becomes an "unwilling host" to the spatial order of the yurt finds direct support in the text, although Claude did not know this and said plainly that the thesis "stands or falls on your texts".

Interpretation in the light of other studies. These findings agree with the limitations that the review of LLMs in literary studies attributes to shallow reasoning and data bias [1], and they are consistent with evidence that models write fluent interpretations [2, 3]. The fluency is there; grounding in the text is not. The mechanism of cognitive offloading [4, 5] is visible in the very shape of the February dialogue. Every answer ended by offering the next ready-made product, a

paragraph, a table or a set of quotations, rather than a question for the student. ChatGPT and Claude behaved more like a tutor with guardrails, the condition that protected learning in the experimental evidence [5]. Cultural bias [13, 14] was milder than we expected. The cultural vocabulary was mostly correct, and the errors lay not in the words but in knowledge of a particular Kazakh text. For academic integrity the practical lesson is that a log of the student's dialogue with the model tells a teacher more than an AI-text detector, and detectors are also known to misjudge writing by non-native speakers [20]. Assessment should look at the process as well as the product [21].

Limitations. We examined one dialogue scenario, with one run per model. Model output is stochastic, and a second run could produce different figures. Access conditions differed (Table 1): the February Gem ran with its own system instructions, ChatGPT and Gemini-09 were used without signing in, and Claude was used in an account whose personal preferences ask the model to test the user's claims critically. That setting probably increased Claude's readiness to challenge the framing, so its score on this indicator should be read as an upper estimate. The coding was done by one researcher, and inter-rater agreement was not measured. The electronic text of the novel was not collated with a printed edition. The results should therefore be treated as grounded hypotheses for a study with students.

Conclusion

The hypothesis was confirmed. The models are comparable in what might be called work before the text: they suggest directions, analytical categories, a structure for the paper and cultural vocabulary. Their epistemic behaviour differs considerably. In both sessions, eight months apart, Gemini accepted the user's framing, did not ask which texts the user had read, and gave details about the novel without sources. Some of those details turned out to be good genre guesses and some were invented, and only reading the primary source told the two apart. ChatGPT avoided claims about the novel's content and directed the user to the text. Claude drew on the sources it found, challenged the framing and made one factual error.

For the theory and practice of educational measurement, the study proposes a set of operational indicators of the epistemic quality of an AI response (Table 2): the share of claims with a source, the share of claims confirmed by the primary text, how often the model refers the user to a corpus, and how often it challenges the user's framing. Teachers can use these indicators when choosing a tool, and researchers when planning an experiment.

Our practical recommendations fall into three stages. Before the text, AI is acceptable as a source of directions, questions to ask of an interior and cultural terms, provided these are checked against a dictionary. Instead of the text, that is, when the model makes claims about the content of a work, its answer should not be accepted without a page reference and a quotation that the student has found. After the text, the model is useful as an opponent that looks for objections to the student's draft but does not write the analysis. The material of this study yields a ready seminar task: students receive Gemini's February answer about Zharmak, check each of its eleven claims against the novel, and then rewrite the model's "ready-made paragraph" so that every point rests on a quotation. The task trains close reading and a critical attitude to AI at the same time, in line with the goals of the national AI Concept and the AI-Sana programme [7, 8] and with the principle of human agency in the UNESCO guidance [22].

The next step is a study with groups of philology students in which work across the three stages is compared with traditional analysis and the quality of students' interpretations of interiors is measured with a rubric of tested inter-rater reliability.

References

1. Mozaffari N., Mehrangmarani B., Khorramrouz A. Applications, performance, and research gaps of large language models in literary studies: a scoping review // *Human Behavior and Emerging Technologies*. – 2026. – Vol. 2026. – Art. 8695447. – <https://doi.org/10.1155/hbe2/8695447>
2. Ichien N., Stamenković D., Holyoak K. J. Large language model displays emergent ability to interpret novel literary metaphors // *Metaphor and Symbol*. – 2024. – Vol. 39, № 4. – P. 296–309. – <https://doi.org/10.1080/10926488.2024.2380348>
3. Bellot A. R., Gutiérrez-Colón Plana M., Baran K. A. Redefining literature education: the role of ChatGPT in undergraduate courses // *International Journal of Artificial Intelligence in Education*. – 2025. – Vol. 35, № 5. – P. 3185–3201. – <https://doi.org/10.1007/s40593-025-00497-3>
4. Fan Y., Tang L., Le H., Shen K., Tan S., Zhao Y., Shen Y., Li X., Gašević D. Beware of metacognitive laziness: effects of generative artificial intelligence on learning motivation, processes, and performance // *British Journal of Educational Technology*. – 2025. – Vol. 56, № 2. – P. 489–530. – <https://doi.org/10.1111/bjet.13544>
5. Bastani H., Bastani O., Sungu A., Ge H., Kabakcı Ö., Mariman R. Generative AI without guardrails can harm learning: evidence from high school mathematics // *Proceedings of the National Academy of Sciences of the USA*. – 2025. – Vol. 122, № 26. – Art. e2422633122. – <https://doi.org/10.1073/pnas.2422633122>
6. Sharma M., Tong M., Korbak T. [et al.] Towards understanding sycophancy in language models // *The Twelfth International Conference on Learning Representations (ICLR 2024)*. – 2024. – <https://arxiv.org/abs/2310.13548>.
7. Об утверждении Концепции развития искусственного интеллекта на 2024–2029 годы: постановление Правительства Республики Казахстан от 24 июля 2024 года № 592 // ИПС «Әділет». – <https://adilet.zan.kz/rus/docs/P2400000592>.
8. Об утверждении Паспорта проекта инновационной акселерационной программы AI-Sana: приказ Министра науки и высшего образования Республики Казахстан от 3 июня 2025 года № 290 // ИС «Параграф». – URL: https://prg.kz/document/?doc_id=31706872.
9. Об искусственном интеллекте: Закон Республики Казахстан от 17 ноября 2025 года № 230-VIII // ИПС «Әділет». – URL: <https://adilet.zan.kz/rus/docs/Z2500000230>.
10. Tazhibayeva S., Kanatkerey A., Mamayeva G., Beisenuly Z., Buribayeva M., Kemberbay R. The impact of artificial intelligence applications on linguistic and literary connections among Turkic peoples based on ethnocultural codes // *Journal of Ethnic and Cultural Studies*. – 2026. – Vol. 13, № 3. – P. 207–228. – <https://doi.org/10.66815/ejecs/3042>
11. Bekturova M., Tulepova S., Zhaitapova A. Predicting Kazakhstani TEFL students' continuance intention towards using ChatGPT in academic writing // *Education and Information Technologies*. – 2025. – Vol. 30, № 9. – P. 11889–11909. – <https://doi.org/10.1007/s10639-024-13306-6>
12. Potluri R. M., Jumasseitova A. K., Mukashev Y. Artificial intelligence in Kazakhstan higher education: computing students' perceptions, engagement, and skill development // *International Journal of Educational Research Open*. – 2026. – Vol. 10. – Art. 100573. – <https://doi.org/10.1016/j.ijedro.2026.100573>
13. Tao Y., Viberg O., Baker R. S., Kizilcec R. F. Cultural bias and cultural alignment of large language models // *PNAS Nexus*. – 2024. – Vol. 3, № 9. – Art. pgae346. – <https://doi.org/10.1093/pnasnexus/pgae346>
14. Togmanov M., Mukhituly N., Turmakhan D., Mansurov J., Goloburda M., Sakip A., Xie Z., Wang Y., Syzdykov B., Laiyk N., Aji A. F., Kochmar E., Nakov P., Koto F. KazMMLU: evaluating language models on Kazakh, Russian, and regional knowledge of Kazakhstan // *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Vol. 1: Long Papers)*. – Vienna: ACL, 2025. – P. 14403–14416. – <https://doi.org/10.18653/v1/2025.acl-long.701>
15. Лотман Ю. М. Структура художественного текста. – М.: Искусство, 1970. – 384 с.
16. Башляр Г. Избранное: Поэтика пространства / пер. с франц. – М.: РОССПЭН, 2004. – 376 с.
17. Чудаков А. П. Слово – вещь – мир. От Пушкина до Толстого: очерки поэтики русских классиков. – М.: Современный писатель, 1992. – 320 с.
18. Әбікенұлы С. Мұхтар-Иманның мұны немесе «Жармақ» // *Заң газеті*. – 2010. – 5 қаңтар. – <https://abai.kz/post/3006>
19. Мағауин М. Жармақ: роман // *Zharar.com*. – <https://zharar.com/index.php?action=item&do=shorttexts&id=20339>

20. Liang W., Yuksekgonul M., Mao Y., Wu E., Zou J. GPT detectors are biased against non-native English writers // Patterns. – 2023. – Vol. 4, № 7. – Art. 100779. – <https://doi.org/10.1016/j.patter.2023.100779>

21. MLA-CCCC Joint Task Force on Writing and AI. Working paper: overview of the issues, statement of principles, and recommendations. – Modern Language Association; Conference on College Composition and Communication, 2023. – URL: <https://aiandwriting.hcommons.org/working-paper-1>

22. Miao F., Holmes W. Guidance for generative AI in education and research. – Paris: UNESCO, 2023. – 44 p. – URL: <https://unesdoc.unesco.org/ark:/48223/pf0000386693>

М.Е. Оспаналы, А.Қ. Қалиева, Н. Темур

ӘДЕБИЕТТІ ОҚЫТУДАҒЫ ЖИ ЖАУАПТАРЫН БАҒАЛАУ: ҚАЗАҚ ПРОЗАСЫНДАҒЫ ӘДЕБИ ИНТЕРЬЕР

Мақалада генеративті тілдік модельдердің филолог-студенттердің көркем мәтіндегі интерьер поэтикасын саралау және түсіндіру құзыретін қалыптастырудағы әдістемелік әлеуеті зерделенеді. Зерттеудің эмпирикалық негізін 2026 жылдың ақпан айында Google Gemini жүйесімен жүргізілген алғашқы диалог пен араға сегіз ай салып Gemini, ChatGPT және Claude модельдеріне қайта қойылған сәйкес сауалдар нәтижесі құрады. Модельдер генерациялаған жауаптар алдын ала белгіленген нақты өлшемдер бойынша салғастырылып, олардың М. Мағауиннің «Жармақ» романындағы интерьерге қатысты пайымдаулары шығарманың түпнұсқа мәтінімен және әдеби-сыни рецензиямен салыстырыла тексерілді. Талдау барысы қарастырылған барлық модельдердің әдіснамалық бағдар беру деңгейінде пайдалы болғанын көрсетті. Жасанды интеллект ұсынған талдау сұрақтары Ю. М. Лотман, Г. Башляр және А. П. Чудаковтың көркем кеңістік туралы тұжырымдамаларымен үндестік тауып, қазақ дәстүрлі баспанасының ұғымдық аппараты дұрыс қолданылды. Алайда нақты көркем мәтінді терең талдауда модельдердің сенімділік деңгейінің шектеулі екені анықталды. Gemini ұсынған 11 тұжырымның тек 2 тұжырымы толық, 6 тұжырымы ішінара расталды. ChatGPT қолданушыны бастапқы мәтінге қайта бағыттаумен шектелді, ал Claude веб-дереккөздерге сүйенді. Романның мәнді кеңістік детальдары модельдер арқылы емес, тек мәтінді мұқият оқу тәсілімен ғана айқындалды. Зерттеу нәтижесінде жоғары оқу орындарының филологиялық курстарына арналған жасанды интеллектпен жұмыс істеудің үш кезеңді оқыту үлгісі мен студенттердің сыни талдау дағдысын шыңдайтын семинар тапсырмалары әзірленді.

Түйін сөздер: генеративті жасанды интеллект; үлкен тілдік модельдер; әдебиетті оқыту; әдеби интерьер; көркем кеңістік; қазақ әдебиеті.

М.Е. Оспаналы, А.Қ. Қалиева, Н. Темур

ОЦЕНКА ОТВЕТОВ ИИ В ПРЕПОДАВАНИИ ЛИТЕРАТУРЫ: ЛИТЕРАТУРНЫЙ ИНТЕРЬЕР В КАЗАХСКОЙ ПРОЗЕ

В статье изучается методический потенциал генеративных языковых моделей в формировании у студентов-филологов умения анализировать и интерпретировать поэтику интерьера в художественном тексте. Эмпирическую основу исследования составили первоначальный диалог с системой Google Gemini, проведённый в феврале 2026 г., и ответы на те же запросы, повторно заданные спустя восемь месяцев моделям Gemini, ChatGPT и Claude. Ответы моделей сопоставлены по заранее установленным критериям, а их суждения об интерьере в романе М. Мағауина «Жармақ» проверены по тексту произведения и литературно-критической рецензии. Анализ показал, что все рассмотренные модели оказались полезны на уровне методологических ориентиров. Предложенные ИИ вопросы для анализа согласуются с концепциями художественного пространства Ю. М. Лотмана, Г. Башляра и А. П. Чудакова, а понятийный аппарат традиционного казахского жилища использован корректно. Вместе с тем при углубленном анализе конкретного художественного текста надежность моделей оказалась ограниченной. Из 11 суждений Gemini полностью подтвердились лишь 2, частично 6. ChatGPT ограничился тем, что возвращал пользователя к исходному тексту, а Claude

опирался на веб-источники, и из 8 его утверждений подтвердились 5. При этом ни одна модель не привела конкретных цитат из романа. Значимые пространственные детали романа были выявлены не с помощью моделей, а только путем внимательного чтения текста. По итогам исследования разработаны трехэтапная модель обучения работе с искусственным интеллектом для филологических курсов вузов и семинарские задания, развивающие у студентов навыки критического анализа.

Ключевые слова: генеративный искусственный интеллект; большие языковые модели; преподавание литературы; литературный интерьер; художественное пространство; казахская литература.

References

1. Mozaffari, N., Mehrangmarani, B., & Khorramrouz, A. (2026). Applications, performance, and research gaps of large language models in literary studies: A scoping review. *Human Behavior and Emerging Technologies*, 2026, Article 8695447. <https://doi.org/10.1155/hbe2/8695447>
2. Ichien, N., Stamenković, D., & Holyoak, K. J. (2024). Large language model displays emergent ability to interpret novel literary metaphors. *Metaphor and Symbol*, 39(4), 296-309. <https://doi.org/10.1080/10926488.2024.2380348>
3. Bellot, A. R., Gutiérrez-Colón Plana, M., & Baran, K. A. (2025). Redefining literature education: The role of ChatGPT in undergraduate courses. *International Journal of Artificial Intelligence in Education*, 35(5), 3185–3201. <https://doi.org/10.1007/s40593-025-00497-3>
4. Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*, 56(2), 489–530. <https://doi.org/10.1111/bjet.13544>
5. Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences*, 122(26), Article e2422633122. <https://doi.org/10.1073/pnas.2422633122>
6. Sharma, M., Tong, M., Korbak, T., Duvenaud, D., Askill, A., Bowman, S. R., Cheng, N., Durmus, E., Hatfield-Dodds, Z., Johnston, S. R., Kravec, S., Maxwell, T., McCandlish, S., Ndousse, K., Rausch, O., Schiefer, N., Yan, D., Zhang, M., & Perez, E. (2024). Towards understanding sycophancy in language models. In *The Twelfth International Conference on Learning Representations (ICLR 2024)*. <https://arxiv.org/abs/2310.13548>
7. Pravitel'stvo Respubliki Kazakhstan. (2024). *Ob utverzhdenii Kontseptsii razvitiya iskusstvennogo intellekta na 2024–2029 gody* [On approval of the Concept for the development of artificial intelligence for 2024–2029] (Postanovlenie No. 592 ot 24 iyulya 2024 goda). <https://adilet.zan.kz/rus/docs/P2400000592>
8. Ministerstvo nauki i vysshego obrazovaniya Respubliki Kazakhstan. (2025). *Ob utverzhdenii Pasporta proekta innovatsionnoy akseleratsionnoy programmy AI-Sana* [On approval of the project passport of the AI-Sana innovative acceleration programme] (Priказ No. 290 ot 3 iyunya 2025 goda). https://prg.kz/document/?doc_id=31706872
9. Respublika Kazakhstan. (2025). *Ob iskusstvennom intellekte* [On artificial intelligence] (Zakon No. 230-VIII ot 17 noyabrya 2025 goda). <https://adilet.zan.kz/rus/docs/Z2500000230>
10. Tazhibayeva, S., Kanatkerey, A., Mamayeva, G., Beisenuly, Z., Buribayeva, M., & Kemberbay, R. (2026). The impact of artificial intelligence applications on linguistic and literary connections among Turkic peoples based on ethnocultural codes. *Journal of Ethnic and Cultural Studies*, 13(3), 207–228. <https://doi.org/10.66815/ejecs/3042>
11. Bekturova, M., Tulepova, S., & Zhaitapova, A. (2025). Predicting Kazakhstani TEFL students' continuance intention towards using ChatGPT in academic writing. *Education and Information Technologies*, 30(9), 11889–11909. <https://doi.org/10.1007/s10639-024-13306-6>
12. Potluri, R. M., Jumasseitova, A. K., & Mukashev, Y. (2026). Artificial intelligence in Kazakhstan higher education: Computing students' perceptions, engagement, and skill development. *International Journal of Educational Research Open*, 10, Article 100573. <https://doi.org/10.1016/j.ijedro.2026.100573>
13. Tao, Y., Viberg, O., Baker, R. S., & Kizilcec, R. F. (2024). Cultural bias and cultural alignment of large language models. *PNAS Nexus*, 3(9), Article pgae346. <https://doi.org/10.1093/pnasnexus/pgae346>
14. Togmanov, M., Mukhituly, N., Turmakhan, D., Mansurov, J., Goloburda, M., Sakip, A., Xie, Z., Wang, Y., Syzdykov, B., Laiyk, N., Aji, A. F., Kochmar, E., Nakov, P., & Koto, F. (2025). KazMMLU:

Evaluating language models on Kazakh, Russian, and regional knowledge of Kazakhstan. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 14403–14416). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2025.acl-long.701>

15. Lotman, Yu. M. (1970). *Struktura khudozhestvennogo teksta* [The structure of the artistic text]. Iskustvo.

16. Bashlyar, G. (2004). *Izbrannoe: Poetika prostranstva* [Selected works: The poetics of space]. ROSSPEN. (Original work published 1957)

17. Chudakov, A. P. (1992). *Slovo – veshch' – mir. Ot Pushkina do Tolstogo: ocherki poetiki russkikh klassikov* [Word – thing – world. From Pushkin to Tolstoy: Essays on the poetics of Russian classics]. Sovremennyy pisatel'.

18. Abikenuly, S. (2010, January 5). Mukhtar-Imannyn muny nemese «Zharmak» [Mukhtar-Iman's sorrow, or «Zharmak»]. *Zan gazeti*. <https://abai.kz/post/3006>

19. Magauin, M. (2007). *Zharmak: roman* [Zharmak: A novel]. Zharar.com. <https://zharar.com/index.php?action=item&do=shorttexts&id=20339>

20. Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. *Patterns*, 4(7), Article 100779. <https://doi.org/10.1016/j.patter.2023.100779>

21. MLA-CCCC Joint Task Force on Writing and AI. (2023). *MLA-CCCC Joint Task Force on Writing and AI working paper: Overview of the issues, statement of principles, and recommendations*. Modern Language Association; Conference on College Composition and Communication. <https://aiandwriting.hcommons.org/working-paper-1/>

22. Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO. <https://doi.org/10.54675/EWZM9535>

Information about authors:

Ospanaly Madina Eldoskyzy (corresponding author) - doctoral student, Al-Farabi Kazakh National University, Almaty, Kazakhstan, e-mail: ospanaly.kz@gmail.com

Kalievа Almira - Candidate of Philological Sciences, M. Auezov Institute of Literature and Art, Almaty, Kazakhstan, e-mail: almira_kalievа.8@mail.ru

Temur Nezir - Doctor, Professor, Gazi University, Ankara, Turkey, e-mail: ntemur@gazi.edu.tr

Авторлар туралы мәлімет:

Оспаналы Мадина Елдоскызы (автор-корреспондент) - докторант, әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан, e-mail: ospanaly.kz@gmail.com

Қалиева Альмира Қайыртайқызы - филология ғылымдарының кандидаты, М. Әуезов атындағы Әдебиет және өнер институты, Алматы, Қазақстан, e-mail: almira_kalievа.8@mail.ru

Темур Незир - доктор, профессор, Гази университеті, Анкара, Түркия, e-mail: ntemur@gazi.edu.tr

Сведения об авторах:

Оспаналы Мадина Елдоскызы (автор-корреспондент) - докторант, Казахский национальный университет имени аль-Фараби, Алматы, Казахстан, e-mail: ospanaly.kz@gmail.com

Калиева Альмира Кайыртаевна - кандидат филологических наук, Институт литературы и искусства имени М. Ауэзова, Алматы, Казахстан, e-mail: almira_kalievа.8@mail.ru

Темур Незир - доктор, профессор, Университет Гази, Анкара, Турция, e-mail: ntemur@gazi.edu.tr