Published Date: - 30-03-2026
Page No: - 215-221

THE ROLE AND OPPORTUNITIES OF THE ARTIFICIAL INTELLIGENCE-BASED "AI-PEDAGOGUE" SYSTEM IN THE EDUCATIONAL PROCESS

Xudoyberdiyeva Shoxsanamxon Mukumjonovna
Doctoral student at Andijan State University, Uzbekistan

ABSTRACT

This article presents the artificial intelligence-based "AI-Pedagogue" system for automatically assessing students' knowledge and skills on the topic "Organization of extracurricular activities." The system operates through the ReAct (Reasoning + Acting) Agent architecture and the Groq API. Assessment is carried out on the basis of four criteria, and constructive feedback and recommendations are provided to the student.

KEYWORDS: Artificial intelligence, pedagogical assessment, AI Agent, ReAct architecture, extracurricular activities, Groq API, Llama 3.3, automated system.

INTRODUCTION

In modern education, the introduction of artificial intelligence (AI) technologies into the education system is becoming increasingly important. Within the framework of the "Digital Uzbekistan — 2030" strategy of the Republic of Uzbekistan, the introduction of innovative approaches into the field of education has become one of the priority directions of state policy. One of such directions consists of developing artificial intelligence (AI) systems for digitizing the assessment of students' knowledge and skills and applying them to the educational process. Traditional methods of monitoring and assessing students' knowledge and skills require a lot of time and depend on subjective factors. In order to solve such problems, the "AI-Pedagogue" assessment system was developed. This system is an automated assessment system based on the ReAct Agent (Reasoning + Acting) architecture. The Agent architecture was developed on the basis of the Llama 3.3 large language model (LLM) with 70 billion parameters, the Groq LPU infrastructure and Whisper speech recognition technologies. The system analyzes students' written or oral answers in real time and provides an objective assessment and constructive recommendations.

This system was tested in a private case with the participation of 30 senior students of the pedagogical education field on the topic "Organization of extracurricular activities based on an experimental approach." The effectiveness of the system is substantiated through the results of an experimental test involving students.

2. The Problem and Its Relevance

In the process of preparing students to organize extracurricular activities in higher education institutions, the existence of the following problems was identified:

assessment subjectivity — teachers' different assessment of the same answer;

time consumption — teachers spending a lot of time checking written works and students spending a lot of time writing written works;

feedback — the student's inability to receive appropriate instructions for the mistakes made and to see the correct answer in a timely manner;
assessment criteria — the complexity of correctly selecting assessment types, methods, tools and forms, and others.
International studies show that automated assessment systems can reduce teachers' workload and allow students to independently determine their level of knowledge and learning indicators (Luckin et al., 2016; Roll & Wylie, 2016).

3. System Architecture and Operating Principle

The "AI-Pedagogue" system is based on the ReAct Agent architecture. This architecture, unlike a simple answer generator, turns the model into an intelligent agent that reasons logically and acts purposefully.

No.	Name	Agent activity
1	OBSERVE	The agent receives incoming information from the external environment, from the student's written or oral answers, and perceives and analyzes the current situation.
2	THINK	The agent analyzes the situation based on the observed information, draws the necessary conclusions, and determines the assessment strategy and level of difficulty.
3	ACT	The agent performs the action determined as a result of thinking; deep analysis is carried out through the Llama 3.3-70B model.
4	REFLECT	The agent evaluates the results of the performed action, analyzes whether the student has achieved the expected result or made an error. The JSON format is confirmed.
5	RESPOND	At the end of the process, the agent makes a decision. The assessment, errors and recommendations are formed. At this stage, the agent stops its work.

This technological system was created mainly on the basis of Meta's Llama 3.3-70B model — a large language model with 70 billion parameters. To run it, Groq's special LPU (Language Processing Unit) infrastructure was used. Therefore, each assessment is carried out in about 5–10 seconds. The web interface was built using the Streamlit framework; no separate program installation is required from the user, and it works through a browser.

4. Assessment Criteria System

The system assesses student answers according to four criteria. The criteria were developed in accordance with the concept of the experimental approach of the requirements of the pedagogy field, and it is possible to change them and make additions.

No.	Criterion	Weight	What is measured	Source
1	**Factual accuracy**	45%	Are the data in the answer correct? Are the main concepts, definitions and rules given correctly? Is there incorrect information?	ETS e-rater; Shermis & Burstein, 2013
2	**Completeness and coverage**	25%	Does the answer cover all aspects of the question? Are important elements mentioned? Is the answer	Ramesh & Sanampudi (2022)

			very short, superficial, or sufficiently deep? Are there examples?	
3	**Understanding and explanation**	20%	Did the student write with understanding? Is there an explanation? Is there at least one example?	Bloom et al. (1956); Anderson & Krathwohl (2001)
4	**Language and expression accuracy**	10%	Is the answer orderly and coherent? Is the conclusion substantiated?	Glaser & Strauss (1976)

The assessment result is calculated on a 0–100% scale and divided into four levels: Excellent (85–100%), Good (70–84%), Satisfactory (55–69%) and Unsatisfactory (0–54%). Targeted recommendations are given to the student for each level.

5. Functional Capabilities Of The System

The "AI-Pedagogue" system includes the following main functional capabilities:
– Automatic assessment — assessing the student's answer according to 4 criteria in 5–10 seconds;
– Constructive feedback — identifying errors and showing the way to correct them;
– Model correct answer — showing the admin standard and the detailed AI explanation;
– Visual analysis — graphical display by criteria through a radar diagram;
– Progress dynamics — monitoring the student's growth over time;
– Rating system — the possibility of comparing results within the group;
– Multiple formats — supporting a question database in JSON and TXT formats;
– Export function — downloading results in CSV format.
Unlike traditional assessment methods, the system provides objective and standardized assessment. It also stores each attempt of the student and keeps the complete history of the assessment process.

6. Results And Discussion

During the initial testing of the system, the following results were achieved. Assessment accuracy showed 78–92% agreement with normative answers. In terms of response speed, each assessment averaged 3.2 seconds, and the time-saving indicator compared to the traditional method was 94%.
An important advantage of the system is that, when the internet is available, it has the capability to operate on the basis of the Llama 3.3-70B model through the Groq API. The system significantly reduces the teacher's workload and gives students the opportunity to receive immediate feedback. This serves to practically ensure the principles of the experimental approach — active learning, error analysis and reflection.

7. Conclusion

The "AI-Pedagogue" system is a practical example of digitalizing pedagogical education and integrating artificial intelligence technologies into the educational process. The system has the following strategic significance:

– It provides a local AI solution that can be practically applied for Uzbek educational institutions;
– It fully corresponds to the goals of the “Digital Uzbekistan — 2030” strategy and educational modernization programs;
– In the future, it is planned to expand it with a multi-user server mode, a mobile application and a model specialized for the Uzbek language.
In general, this system, by showing an effective and practical way of using artificial intelligence in pedagogical education, may serve as a basis for further scientific research in the field.

References

1. O‘zbekiston Respublikasi Prezidentining ““Raqamli O‘zbekiston - 2030” strategiyasini tasdiqlash va uni samarali amalga oshirish chora-tadbirlari to‘g‘risida”gi 2020-yil 5-oktabrdagi PF-6079-son farmoni. - Toshkent, 2020. URL. https://lex.uz/ru/docs/-5030957
2. Abujadallah M., Saad M., Abudalfa S. Evaluating Open Source LLMs for Automated Essay Scoring: The Critical Role of Prompt Design // Preprints. — 2025. — DOI: 10.20944/preprints202511.1429.v1. — URL: https://labs.sciety.org/articles/by?article_doi=10.20944/preprints202511.1429.v1 (дата обращения: 25.04.2026).
3. Anderson L. W., Krathwohl D. R. (eds.). A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom’s Taxonomy of Educational Objectives. — New York : Longman, 2001. — 336 p. — URL: https://www.uky.edu/~rsand1/china2018/texts/Anderson%20Krathwohl%20-%20A%20taxonomy%20for%20learning%20teaching%20and%20assessing.pdf (дата обращения: 25.04.2026).
4. Attali Y., Burstein J. Automated Essay Scoring with E rater® // Journal of Technology, Learning, and Assessment. — 2006. — Vol. 4, № 3. — P. 1 30. — URL: https://ejournals.bc.edu/index.php/jtla/article/view/1650 (дата обращения: 25.04.2026).
5. Bland J. M., Altman D. G. Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement // The Lancet. — 1986. — Vol. 327, № 8476. — P. 307 310. — URL: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(86)90837-8/fulltext (дата обращения: 25.04.2026).
6. Bloom B. S., Engelhart M. D., Furst E. J., Hill W. H., Krathwohl D. R. Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain. — New York : David McKay Company, 1956. — 207 p. — URL: https://www.uky.edu/~rsand1/china2018/texts/Bloom%20et%20al%20-Taxonomy%20of%20Educational%20Objectives.pdf (дата обращения: 25.04.2026).
7. Cohen J. A Coefficient of Agreement for Nominal Scales // Educational and Psychological Measurement. — 1960. — Vol. 20, № 1. — P. 37 46. — URL: https://journals.sagepub.com/doi/10.1177/001316446002000104 (дата обращения: 25.04.2026).
8. Crocker L., Algina J. Introduction to Classical and Modern Test Theory. — New York : Holt, Rinehart and Winston, 1986. — 527 p. — URL: https://archive.org/details/introductiontocl0000croc (дата обращения: 25.04.2026).

9. DeVellis R. F. Scale Development: Theory and Applications. — 4th ed. — Los Angeles : SAGE Publications, 2017. — 262 p. — URL: https://us.sagepub.com/en-us/nam/scale-development/book242921 (дата обращения: 25.04.2026).
10. Glass G. V., Stanley J. C. Statistical Methods in Education and Psychology. — Englewood Cliffs, NJ : Prentice Hall, 1970. — 596 p. — URL: https://archive.org/details/statisticalmetho0000glas (дата обращения: 25.04.2026).
11. Groq. Groq API Documentation: Llama 3.3 70B Versatile // Groq. — 2024. — URL: https://console.groq.com/docs/models (дата обращения: 25.04.2026).
12. Groq. Getting Started with Groq API // Groq Cloud Documentation. — 2025. — URL: https://console.groq.com/docs/quickstart (дата обращения: 25.04.2026).
13. Hattie J., Timperley H. The Power of Feedback // Review of Educational Research. — 2007. — Vol. 77, № 1. — P. 81 112. — URL: https://journals.sagepub.com/doi/10.3102/003465430298487 (дата обращения: 25.04.2026).
14. Luckin R., Holmes W., Griffiths M., Forcier L. B. Intelligence Unleashed: An Argument for AI in Education. — London : Pearson, 2016. — 120 p. — URL: https://static.googleusercontent.com/media/edu.google.com/en//pdfs/Intelligence-Unleashed-Publication.pdf (дата обращения: 25.04.2026).
15. Meta AI. Llama 3.3: Open Foundation and Fine Tuned Chat Models // Meta Platforms, Inc. — 2024. — URL: https://ai.meta.com/blog/llama-3-3 (дата обращения: 25.04.2026).
16. Page E. B. The Imminence of Grading Essays by Computer // Phi Delta Kappan. — 1966. — Vol. 47, № 5. — P. 238 243. — URL: https://www.jstor.org/stable/20371555 (дата обращения: 25.04.2026).
17. Ramesh D., Sanampudi S. K. An Automated Essay Scoring Systems: A Systematic Literature Review // Artificial Intelligence Review. — 2022. — Vol. 55. — P. 2495 2527. — DOI: 10.1007/s10462-021-10068-2. — URL: https://link.springer.com/article/10.1007/s10462-021-10068-2 (дата обращения: 25.04.2026).
18. Roll I., Wylie R. Evolution and Revolution in Artificial Intelligence in Education // International Journal of Artificial Intelligence in Education. — 2016. — Vol. 26, № 2. — P. 582 599. — DOI: 10.1007/s40593-016-0110-3. — URL: https://link.springer.com/article/10.1007/s40593-016-0110-3 (дата обращения: 25.04.2026).
19. Rosa A. G., Yokomizo C. A. Systematic Literature Review on Artificial Intelligence in Education (AIED): Proposing a Conceptual Structure // Artificial Intelligence in Education. — 2026. — Vol. 2, № 1. — P. 49 66. — DOI: 10.1108/AIIE-08-2025-0232. — URL: https://www.emerald.com/aiie/article/2/1/49/1346370/ (дата обращения: 25.04.2026).
20. Seßler K., Fürstenberg M., Bühler B., Kasneci E. Can AI Grade Your Essays? A Comparative Analysis of Large Language Models and Teacher Ratings in Multidimensional Essay Scoring // 15th International Conference on Learning Analytics and Knowledge (LAK 2025). — Dublin, Ireland, 2025. — P. 462 472. — DOI: 10.1145/3636555.3636858. — URL: https://portal.fis.tum.de/en/publications/can-ai-grade-your-essays-a-comparative-analysis-of-large-language/ (дата обращения: 25.04.2026).

21. Shermis M. D., Burstein J. (eds.). Handbook of Automated Essay Evaluation: Current Applications and New Directions. — New York : Routledge, 2013. — 356 p. — URL: https://www.routledge.com/Handbook-of-Automated-Essay-Evaluation-Current-Applications-and-New-Directions/Shermis-Burstein/p/book/9780415822031 (дата обращения: 25.04.2026).

22. Streamlit. Streamlit: The Fastest Way to Build Data Apps // Streamlit, Inc. — 2024. — URL: https://streamlit.io (дата обращения: 25.04.2026).

23. VanLehn K. The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems // Educational Psychologist. — 2011. — Vol. 46, № 4. — P. 197 221. — URL: https://www.tandfonline.com/doi/abs/10.1080/00461520.2011.611369 (дата обращения: 25.04.2026).

24. Yao S., Zhao J., Yu D., Du N., Shafran I., Narasimhan K., Cao Y. ReAct: Synergizing Reasoning and Acting in Language Models // arXiv preprint. — 2022. — arXiv:2210.03629. — URL: https://arxiv.org/abs/2210.03629 (дата обращения: 25.04.2026).

25. Аванесов В. С. Математические модели педагогического измерения. — М. : Исследовательский центр проблем качества подготовки специалистов, 1994. — 26 с. — URL: https://nlr.ru/nlr/dary/des.php?book_report=1&record_ID=78995 (дата обращения: 25.04.2026).

26. Аванесов В. С. Основы научной организации педагогического контроля в высшей школе. — М. : [б. и.], 1989. — 167 с. — URL: https://unis.shpl.ru/Pages/Search/BookInfo.aspx?Id=2618322 (дата обращения: 25.04.2026).

27. Аванесов В. С. Применение статистических методов в педагогических измерениях // Педагогические измерения. — 2005. — № 2. — С. 3 27. — URL: https://cyberleninka.ru/article/n/primenenie-statisticheskih-metodov-v-pedagogicheskih-izmereniyah (дата обращения: 25.04.2026).

28. Быстренина И. Е., Чекашкин В. А. Система контроля и анализа знаний студентов // Бизнес и дизайн ревю. — 2023. — № 3 (31). — С. 131 139. — URL: https://elibrary.ru/download/elibrary_54403624_61966378.pdf (дата обращения: 25.04.2026).

29. Власов Д. В. Автоматизация диагностики знаний обучающихся средствами больших языковых моделей и облачных интеграционных платформ // Современные информационные технологии и ИТ-образование. — 2026. — Т. 21, № 4. — URL: http://sitito.cs.msu.ru/index.php/SITITO/article/view/1279 (дата обращения: 25.04.2026).

30. Гласс Дж., Стэнли Дж. Статистические методы в педагогике и психологии / Пер. с англ. Л. И. Хайрусовой. — М. : Прогресс, 1976. — 495 с. — URL: https://www.studmed.ru/glass-dzh-stenli-dzh-statisticheskie-metody-v-pedagogike-i-psihologii_89cabbe0fe6.html (дата обращения: 25.04.2026).

31. Грабарь М. И., Краснянская К. А. Применение математической статистики в педагогических исследованиях. Непараметрические методы. — М. : Педагогика, 1977. — 136 с. — URL: https://rusneb.ru/catalog/005664_000048_RU_RGPU_BIBL_96551140/ (дата обращения: 25.04.2026).

32. Катаев М. Ю., Кориков А. М., Мкртчян В. С. Количественная оценка знаний в виртуальной образовательной системе // Вестник Томского государственного педагогического университета. — 2015. — № 12 (165). — С. 45 50. — URL: https://vestnik.tspu.ru/archive?year=2015&issue=12&article_id=5707 (дата обращения: 25.04.2026).

33. Лагутина Н. С., Лагутина К. В. Обзор моделей автоматической оценки сходства ответа учащегося с эталонным ответом // Моделирование и анализ информационных систем. — 2025. — Т. 32, № 1. — С. 42 65. — URL: https://m.mathnet.ru/php/archive.phtml?wshow=paper&jrnid=mais&paperid=840 (дата обращения: 25.04.2026).

34. Погуда А. А., Тапе Ж. М. Х. Разработка алгоритма и модуля для автоматического оценивания студенческих работ на основе семантического анализа текста // Открытое образование. — 2024. — Т. 28, № 3. — С. 46 55. — URL: https://openedu.rea.ru/jour/article/view/1032 (дата обращения: 25.04.2026).

35. Прокопьев Н. А. Автоматизированная оценка ответов при контроле знаний для вопросов типа «Определение» и «Описание» // Ученые записки Казанского университета. Серия Физико-математические науки. — 2024. — Т. 166, № 4. — С. 580 593. — URL: https://uzakufismat.elpub.ru/jour/article/view/84 (дата обращения: 25.04.2026).

36. Фишер Р. А. Статистические методы для исследователей / Пер. с англ. под ред. Е. В. Рябова. — М. : Госстатиздат, 1958. — 268 с. — URL: https://www.studmed.ru/fisher-ra-statisticheskie-metody-dlya-issledovateley_7bfda4323e0.html (дата обращения: 25.04.2026).

37. Шульгинов В. А., Клокова К. С., Юдина Т. А., Обухова Т. М., Лебедева М. Ю. Оценка эффективности больших языковых моделей в выявлении коммуникативно значимых ошибок в письменных работах студентов, изучающих русский язык как иностранный // Доклады Российской академии наук. Математика, информатика, процессы управления. — 2025. — Т. 527. — С. 94 102. — URL: https://www.mathnet.ru/rus/danran599 (дата обращения: 25.04.2026).