IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 15, No. 4, August 2026, pp. 3068~3080
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i4.pp3068-3080

Integration of artificial intelligence technology in learning evaluation using the ADDIE method

Adip Wahyudi[1], Slamet Arifin[2], Mukhammad Solikhin[3], Alfi Sahrina[1], Feri Fahrian Maulana[4], Mohammad Firzon Ainur Roziqin[5], Maria Emerita Indraningrum Shrestha[6]
[1]Department of Geography, Faculty of Social Sciences, Universitas Negeri Malang, Malang, Indonesia
[2]Teacher Professional Education Program, Graduate School, Universitas Negeri Malang, Malang, Indonesia
[3]Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Malang, Malang, Indonesia
[4]Environmental Resource Management and Development Program, Graduate School, Universitas Brawijaya, Malang, Indonesia
[5]Department of Electrical Engineering and Informatics, Faculty of Engineering, Universitas Negeri Malang, Malang, Indonesia
[6]The British College, Kathmandu, Nepal

Article Info

Article history:

Received Jul 4, 2025
Revised May 18, 2026
Accepted Jun 19, 2026

Keywords:

ADDIE
Artificial intelligence
Chatbot
Learning evaluation
Technology

ABSTRACT

Artificial intelligence offers significant potential to enhance learning evaluation, yet implementing fair and objective assessments remains a challenge. This study developed an artificial intelligence-based learning evaluation system to improve assessment inclusivity and efficiency. Employing the research and development method with the analysis, design, development, implementation, and evaluation (ADDIE) model, the system features a virtual assistant chatbot and an automated grading tool artificial intelligence assessment. The product underwent validation by media and material experts, followed by field trials with 20 secondary school teachers. The results demonstrated high feasibility, with validation scores of 96% by user interface/user experience (UI/UX) and 91.2% by material experts, and an implementation feasibility score of 94.4%. The system successfully assisted teachers in reducing administrative burdens and providing personalized feedback based on student learning styles. Future developments should expand subject coverage and support diverse assessment formats.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Adip Wahyudi
Department of Geography, Faculty of Social Sciences, Universitas Negeri Malang
Semarang 5 st, Malang 65145, East Java, Indonesia
Email: adip.wahyudi.fis@um.ac.id

1. INTRODUCTION

The development of artificial intelligence technology continues to advance rapidly in 21st century, having a very significant impact on various sectors of life, ranging from industry, health, transportation, education, to entertainment [1]. Artificial intelligence technology has been a catalyst in the digital transformation that is changing the way humans interact with technology and the environment. Artificial intelligence is capable of completing various complex tasks that previously could only be done by humans, such as large-scale data analysis in a short time, facial recognition used for security and authentication, natural language processing (NLP) that enables more intuitive communication with machines, and algorithm-based decision-making that is able to provide solutions accurately and quickly [2]. This progress not only improves work efficiency in various fields, but also creates new opportunities for innovation, accelerates the development of other technologies, and makes a significant contribution to global economic growth [3], [4].

Artificial intelligence in the world of education is applied in smart classrooms to create a learning experience that is more adaptive and responsive to the individual needs of learners [5], [6]. Artificial

intelligence in the education sector continues to develop significantly, used to personalize the learning process, help learners learn at their own pace and style, and ease the administrative burden on teachers, for example with automated grading systems. However, there are concerns that the application of artificial intelligence in education could deepen inequalities, especially if access to technology is uneven [7]. Therefore, it is important to develop an ethical framework that regulates the use of artificial intelligence in education [8]. Principles such as transparency, fairness, and inclusion must be a top priority so that this technology can benefit all groups of society and improve the overall quality of education. For example, in the application of artificial intelligence in the classroom, algorithms must be closely monitored so that there is no bias in the assessment process or recommendations for learning paths for students. In addition, the management of students' personal data needs to be carried out ethically, with priority given to the protection of privacy and individual rights [9]. Artificial intelligence systems designed for education sector must follow international standards in terms of technology security and ethics to avoid unwanted negative impacts [10]. The development of artificial intelligence in education aims not only to improve learning outcomes but also to strengthen the ethical principles that protect the rights of all parties involved.

In recent years the landscape of educational evaluation has been transformed by various artificial intelligence-driven technologies. Systems based on NLP, such as automated essay scoring (AES) and adaptive testing platforms, have been widely adopted to streamline grading processes [11]. However, while general-purpose large language models (LLMs) like ChatGPT demonstrate impressive capabilities in generating feedback, they often lack specific pedagogical context required for secondary education curricula. Most existing tools operate as standalone applications rather than integrated systems designed with structured instructional models like analysis, design, development, implementation, and evaluation (ADDIE) [12]. This gap highlights the need for a specialized evaluation system that combines automated assessment with teacher-validated rubrics.

Based on the results of observations at the Junior High School Laboratory Universitas Negeri Malang (SMP Lab UM) in Malang, Indonesia, it was found that a lack of understanding and skills among teachers in using artificial intelligence technology to accurately assess students' learning progress is a significant problem. The manual evaluation process also leads to delays in providing feedback to students, which of course will hamper a more effective and efficient learning process. Dependence on conventional methods also hampers students' potential for development because it does not provide a quick response regarding the results of their evaluations. Students who need further guidance or improvement in some aspects of learning often have to wait a long time for constructive feedback. In fact, in the digital age, the speed and accuracy of feedback are very important to support students' learning progress according to their individual characteristics. Therefore, it is important for schools to start integrating artificial intelligence technology in learning evaluation in order to provide a more holistic analysis, minimize bias, and create a more adaptive learning process according to the needs of students' learning characteristics.

Many studies have discussed the application of artificial intelligence in the world of education. The integration of artificial intelligence in learning evaluation is still a growing area and requires further exploration. Owan *et al.* [13] highlighted that artificial intelligence can help automate the grading of written assignments and provide instant feedback, but they also acknowledged that the system is still limited in understanding emotional context and subjective nuances in student answers. Meanwhile, Demartini *et al.* [14] shows how machine learning-based artificial intelligence algorithms can identify patterns of student error to provide more personalized learning recommendations. However, this research lacks focus on how artificial intelligence systems can be used to measure non-cognitive abilities, such as collaboration and creativity. Sarker [15] developing an artificial intelligence-based evaluation platform capable of performing multi-dimensional analysis of student performance, but its implementation is limited to a structured digital environment, so it has not been widely adopted in more flexible or hybrid learning contexts. Based on this gap, the novelty that can be offered is to develop an evaluation tool with artificial intelligence integration that not only focuses on evaluating learning outcomes based on cognitive data but is also able to evaluate non-cognitive abilities adaptively and integrate holistic feedback relevant to individual needs. The main objective of integrating artificial intelligence in this learning evaluation is to create a more inclusive, personal, and flexible assessment system so as to improve the quality of education by providing deeper insights into students' learning development, especially in social science learning. The hypothesis of integrating artificial intelligence in learning evaluation is expected to improve the effectiveness of assessment by providing more accurate, personalized, and relevant results compared to traditional evaluation methods. However, alternatively, it is possible that the integration of artificial intelligence does not make a significant difference to the accuracy, personalization, or quality of assessment when compared to traditional methods, especially in certain contexts that require a more complex and contextual evaluation approach.

2. METHOD

Researchers use research and development to produce and evaluate the efficacy of new product [16]. This strategy is very important because it allows researchers to ascertain the needs, design, and create products that are suitable and beneficial to society, especially in the field of education. This research discusses the integration of artificial intelligence in the evaluation of student learning, with teachers as the research subjects or website operators. This research uses the ADDIE instructional design paradigm, which consists of five stages: i) analysis, ii) design, iii) development, iv) implementation, and v) evaluation. Each step plays an important role in ensuring that the final product meets user requirements and effectively achieves its goals. The ADDIE development paradigm is illustrated in Figure 1.

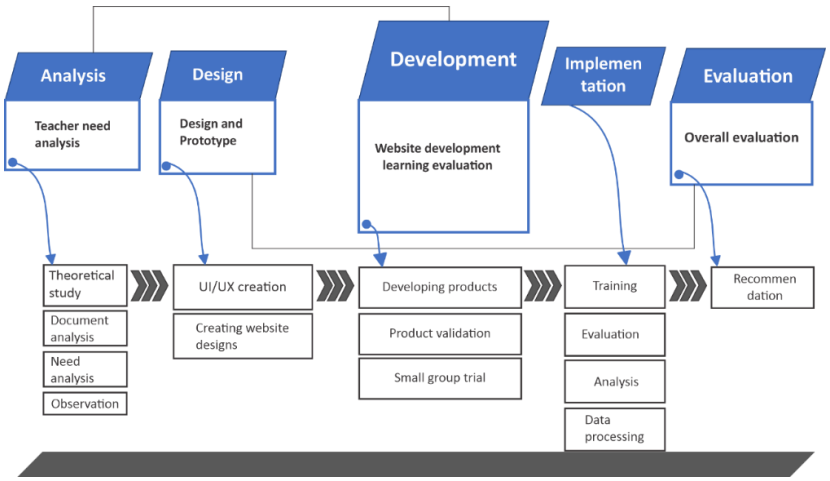

Figure 1. Research flow

The stages of research and development for the integration of artificial intelligence in student learning assessment are described as:

i) Analysis: the strategy for developing an artificial intelligence-based evaluation tool in this study involves an analysis stage that includes needs analysis and instructor needs analysis. The needs analysis attempts to understand the challenges in learning assessment faced by educators and the need to create more effective and efficient evaluation instruments, which includes identifying limitations in traditional assessment methods and prospective artificial intelligence-based solutions. A needs analysis was conducted to ascertain the unique needs of teachers regarding technology-based learning evaluation, ensuring that the resulting system is appropriate, easy to use, and capable of improving the effectiveness of evaluation. The teacher needs analysis focused on digital evaluation related to advances in educational technology, involving teachers as the main subject to understand the requirements and barriers to system implementation. The expected result is the availability of artificial intelligence-based assessment technology that can facilitate effective, efficient, and measurable evaluation of student learning for educators.

ii) Design: the second phase in this development research is design. This phase occurs after the analysis of identified needs, especially those involving the creation of use cases for artificial intelligence-based learning evaluation tools. The use case flow describes the nature of the interaction between the program user and the system. The use case flow describes the interaction between actors, such as educators and learners, and the functions of the artificial intelligence system, including automated feedback, learning outcomes analysis, and personalized evaluation. This image facilitates visualization of the system usage flow and ensures that all user needs are integrated into the design of the artificial intelligence-based evaluation tool.

iii) Development: the third stage in this development research is development. The development of an artificial intelligence-based evaluation tool is carried out in four stages, namely the development of an artificial intelligence-based system, the creation of evaluation instruments, product validation, and product revision. The four stages are explained as:

   – Development of artificial intelligence-based evaluation system: at this stage, dynamic and easy-to-use learning evaluation tools, including artificial intelligence, are designed to meet specific

objectives. The system combines important features such as automated learning outcome analysis, individual feedback, and comprehensive evaluation reports. The evaluation tool consists of a home page, user guide, artificial intelligence-based evaluation function, evaluation report, and user profile page. Each component of this system is designed based on a needs assessment derived from the analysis and design phases.

- Creation of evaluation instruments: the development of the learning evaluation tool was carried out before the data collection stage, using a questionnaire designed to assess the attractiveness, efficacy, and efficiency of the evaluation tool. This instrument has been validated by experts before use. The evaluation instrument grid was created to cover various indications, including usability, the accuracy of artificial intelligence in learning analysis, and relevance to user needs.
- Product validation: the validation phase is carried out after the development of the instrument, involving evaluation experts, education experts, and teachers. This phase includes the evaluation of evaluation tools and artificial intelligence-based practicality. The questionnaire serves as an instrument to evaluate the quality of evaluation tools related to technical aspects and functionality.
- System revision: this phase was carried out after validation and feasibility testing of the evaluation tool. Refinements were implemented in response to input from experts and teachers, which included improving the system's functions, interface, and artificial intelligence assessment algorithms. The end result is an artificial intelligence-based evaluation system that is ready for implementation by teachers to verify its quality and ease of use.

iv) Implementation: stage 4 in the development of this education involves the deployment of an artificial intelligence-based learning assessment system. The implementation stage is used to assess the feasibility of the evaluation instrument when used by educators as users. The trial of the artificial intelligence-based learning evaluation tool will be carried out in two forms: small groups and large groups. The small group trial will consist of four social science educators from the SMP Lab UM in Malang City. In addition, the implementation of the large group will involve 20 teachers from various disciplines. After the trial, the teachers are required to fill out a questionnaire to assess the validity of the evaluation tool and identify issues related to the integration of artificial intelligence, then make revisions to achieve optimal product results.

v) Evaluation: the closing phase in the formulation of the ADDIE model is the evaluation phase. This procedure includes refinement and modification of learning evaluation tools integrated with artificial intelligence. This phase can also occur in the fourth stage, previously known as formative evaluation, which seeks rapid refinement. Currently, evaluation involves refining existing learning evaluation tools to produce evaluation tools that effectively fulfill the desired study objectives. The assessment is carried out after a large group trial conducted by the researcher.

The instrument consists of an open survey that provides qualitative data in the form of criticisms and ideas from expert validators and teachers, as well as a closed questionnaire that provides quantitative data on the evaluation of the resulting artificial intelligence assessment. Table 1 is an observation tool for data collection activities in the field, which is designed to observe and document information directly from the subject or event under study. This instrument is intended to provide a comprehensive picture of the circumstances and conditions in the field, as well as to collect the data needed for further research. In contrast, Table 2 is a needs analysis tool related to the progress of the artificial intelligence assessment. This instrument helps to identify and assess current needs as well as formulate artificial intelligence development strategies in accordance with the desired objectives and arrangements. In addition, Table 3 illustrates aspects of the instrument used in validation by user interface/user experience (UI/UX) specialists. This validation is carried out to ensure that UI and UX designs are up to standard and meet user expectations. Table 4 illustrates the criteria in the validation instrument provided by the academic learning expert. This validation emphasizes pedagogical and methodological elements to guarantee that the proposed tool or system can successfully facilitate the learning process. Table 5 presents the findings from the pilot instrument given to educators at the SMP Lab UM. This study aims to obtain direct input from end users (teachers) regarding the effectiveness and practicality of instruments or systems created for educational purposes in schools. All of these tables are interconnected to ensure that the development and implementation of artificial intelligence in education run effectively and in accordance with user needs and established standards. Researchers use validity analysis and statistical analysis for data examination. Validity analysis uses (1).

$$Average\ Score = \frac{Total\ Score}{Number\ of\ questions} \quad (1)$$

The score obtained from the respondent's assessment is quantitative data, which will later be transformed into qualitative data by referring to the qualification range in Table 6.

Table 1. Observation instruments

No	Observed aspect
1	Teachers' understanding of the use of artificial intelligence in learning evaluation
2	Implementation of artificial intelligence-based learning evaluation by teachers
3	Teachers' enthusiasm for the integration of artificial intelligence in learning Evaluation
4	Teachers' ability to operate artificial intelligence-based evaluation tools
5	Difficulties experienced by teachers in using artificial intelligence-based evaluation tools

Table 2. Instrument grid

No	Indicators	Question topic
1	Teacher analysis	Problems experienced by teachers in using artificial intelligence for evaluation
		Teachers' expectations for the use of artificial intelligence in learning evaluations in the future
2	Needs analysis	Availability of artificial intelligence-based evaluation tools in schools
		Intensity of teacher use of artificial intelligence-based evaluation tools
		Issues related to the use of artificial intelligence tools for learning evaluation
		Teachers' expectations for the development of artificial intelligence-based evaluation tools

Table 3. UI/UX expert instruments

No	Aspect	Question topic
1	Attraction	Intuitive user interface
		Consistency of appearance
		Visual system design
2	Effectiveness	The function of artificial intelligence in evaluation
		Accuracy of evaluation results by artificial intelligence

Table 4. Learning evaluation expert instrument

No	Aspect	Question topic
1	Attraction	Clarity of information presentation evaluation
		Consistency of presentation in all subjects
		Design of universal evaluation results display
2	Effectiveness	Relevance of subject competency evaluation
		Validity and accuracy of artificial intelligence assessment in various subjects

Table 5. Teacher instrument

No	Aspect	Question topic
1	Attraction	Interface design for artificial intelligence systems
		Language clarity in artificial intelligence systems
		Color consistency on the interface
		Image quality and supporting visuals
		Interface layout
2	Effectiveness	Accuracy of artificial intelligence analysis
		Relevance of evaluation of learning
		The accuracy of artificial intelligence feedback
		Scope of evaluation produced
3	Efficiency	Completeness of artificial intelligence evaluation tool features
		Ease of operation evaluation tool
		Usefulness of evaluation results for teachers

Table 6. Criteria and content eligibility [17]

No	Explanation (%)	Class
1	0-20	Very unqualified
2	21-40	Unqualified
3	41-60	Doubtful
4	61-80	Qualified
5	81-100	Very qualified

3. RESULTS AND DISCUSSION

3.1. Analysis

The previous analysis serves as a framework and reference in the implementation of artificial intelligence integration in educational assessment in schools. The analysis includes complexity assessment and evaluation of user needs. The purpose of the complexity analysis is to identify and classify the problems

faced in the educational assessment system in schools related to the implementation of artificial intelligence technology for teachers. This investigation includes an examination of basic and essential competencies related to technology, with an exploration of the principles of artificial intelligence and its application in learning assessment.

Based on the results of interviews and needs assessment questionnaires distributed to social studies teachers in Figure 2, it was found that conventional evaluation methods, such as the use of Excel, result in highly subjective assessments. These manual procedures create a heavy administrative burden and cause educators to feel bored. This indicates an urgent need to integrate artificial intelligence technology to ensure fairer assessments and significantly reduce teachers' administrative tasks.

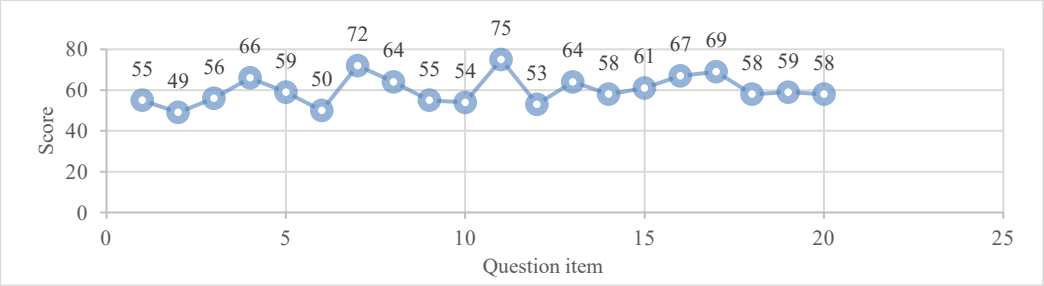

Figure 2. Need analysis result

3.2. Design

Based on the needs identified in the needs analysis stage, the artificial intelligence integration in learning evaluation was developed to meet the needs of a teacher learning evaluation tool. The design process began by selecting an appropriate OpenAI to integrate with the website. A storyboard was then created as a design framework for the chatbot and artificial intelligence assessment features. The storyboard plays an important role in designing the interface, from the home page to the evaluation section. The content is organized by identifying core competencies and basic competencies and determining the type of feature to be used in learning evaluation. The final product developed is a website integrated with artificial intelligence for learning evaluation. The chatbot and artificial intelligence assessment use cases are shown in Figures 3 and 4.

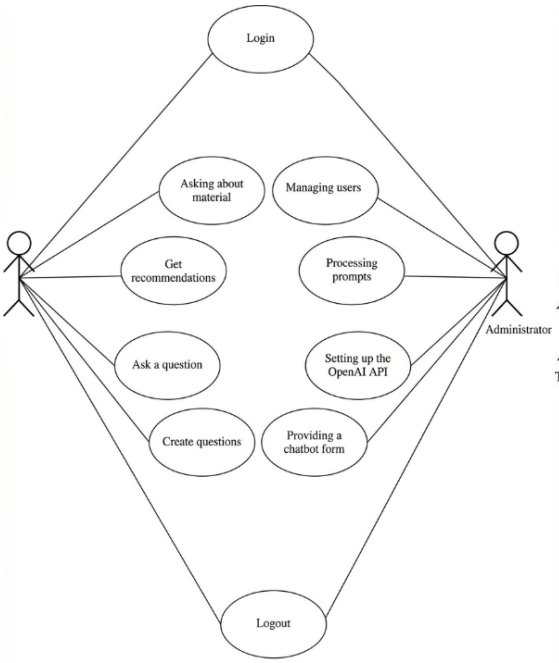

Figure 3. Use case diagram chatbot

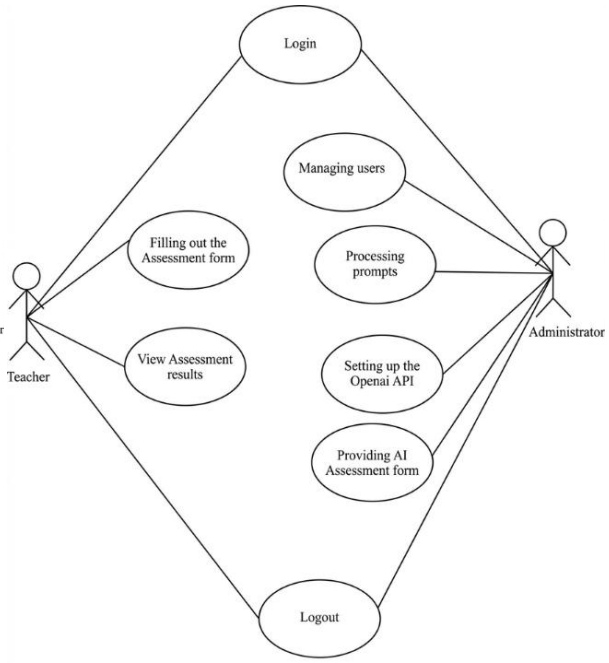

Figure 4. Use case diagram artificial intelligence assessment

3.3. Development

The development of artificial intelligence assessment includes two main components, namely chatbot and artificial intelligence assessment. The chatbot function aims to improve communication between educators and virtual assistants regarding various elements of education, including resource preparation, classroom management, and additional academic questions. Chatbot technology aligns with the concept of artificial intelligence-based conversation agents that increase teacher efficiency by offering direct and relevant access to information [18]. This chatbot, similar to technologies such as ChatGPT 4.0 but adapted for educational purposes, allows educators to quickly obtain guidance or information, speed up the learning planning process, and improve the quality of interaction with students. The idea of artificial intelligence in education supports the function of artificial intelligence as a learning assistant that provides quick feedback on material and improves teaching methodology [19]. This allows educators to align resources more effectively with learner needs and reduce the administrative burden often faced in teaching profession [20]. Simultaneously, the artificial intelligence assessment function concentrates on the automatic evaluation of student assignments, prioritizing the achievement of learning outcomes for specific subjects. This feature aims to increase objectivity and consistency in the assessment process, thus ensuring that student evaluations are based on established learning standards. The theory of automatic assessment supports this idea, where artificial intelligence is used to reduce human bias and increase fairness in evaluation [21]. This technology uses artificial intelligence to correctly assess students' understanding and skills, enabling teachers to provide more appropriate and effective feedback. Figures 5 and 6 show the results of the artificial intelligence assessment product that has been created for the SMP Lab UM.

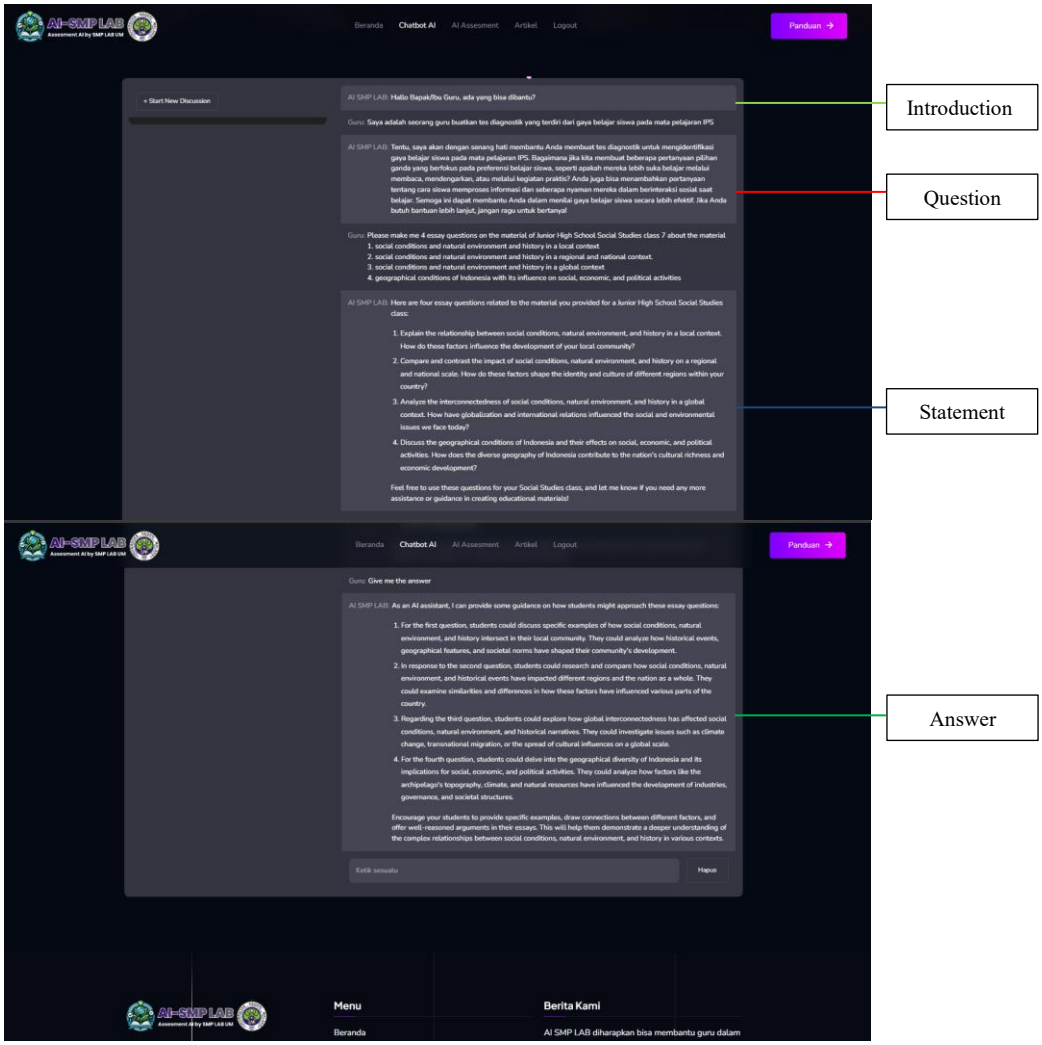

Figure 5. Implementation of chatbot features in social studies material

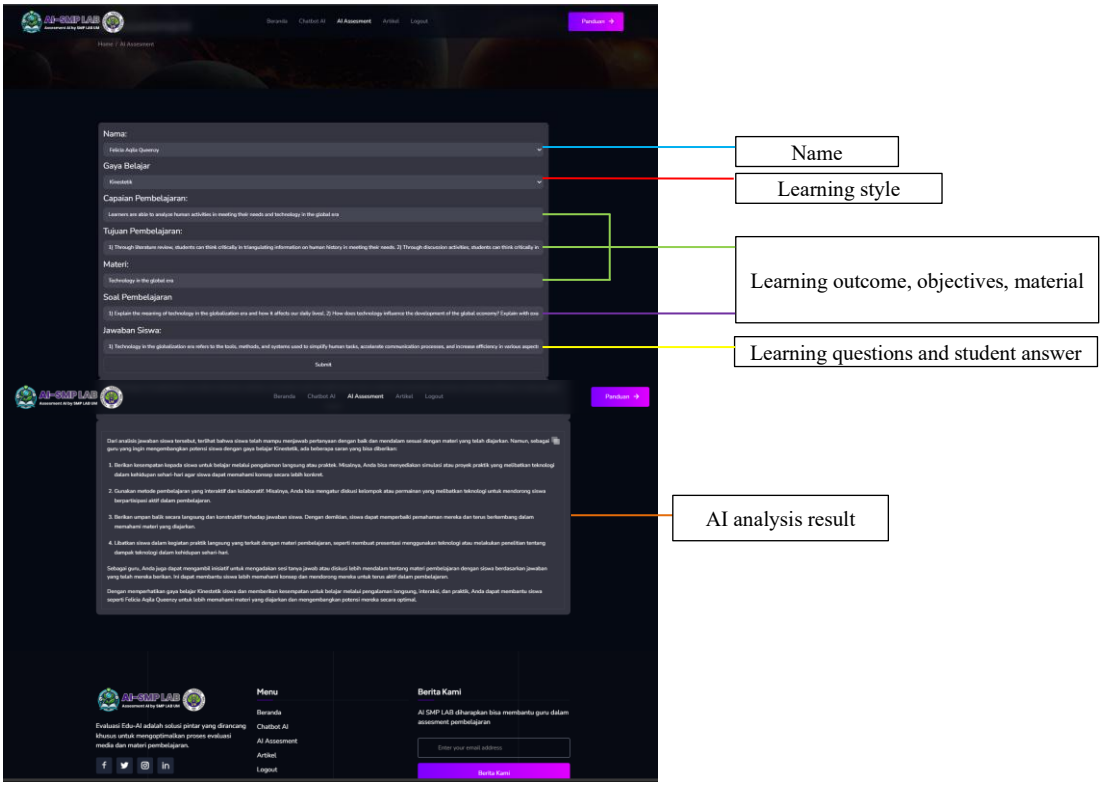

Figure 6. Implementation of AI-SMP Lab assessment

Based on the data presented, the UI/UX validation process was carried out using a questionnaire consisting of 20 statements, along with a comments and suggestions section. The assessment covers several important aspects, such as intuitive user interface, display consistency, system visual design, artificial intelligence function in evaluation, accuracy of evaluation results by artificial intelligence. In Figure 7, the UI/UX aspect obtained a percentage of 96%. This percentage indicates that the artificial intelligence assessment design meets the feasibility criteria with a rating of "very feasible". This indicates that the artificial intelligence assessment has been compiled in detail, clearly, and in accordance with the needs of teachers. The final result of the UI/UX artificial intelligence assessment can help teachers evaluate learning more optimally with a good design. These findings indicate that the UI/UX artificial intelligence assessment is feasible and ready to be piloted with teachers. The eligibility percentage of the learning evaluation validation expert reached 91.2%, as shown in Figure 8. This figure indicates that the artificial intelligence assessment has met the "very feasible" eligibility criteria, making it ready for trial. This artificial intelligence assessment offers automatic evaluation of student assignments, prioritizing the achievement of learning outcomes for specific subjects.

3.4. Implementation

The artificial intelligence assessment was implemented for teachers after being validated by experts in the development stage. This implementation stage involved 20 teachers from various subjects as an experimental group. The assessment was carried out using a questionnaire consisting of 20 statements regarding the components of the artificial intelligence assessment, along with a column for comments and suggestions. The aspects evaluated included attractiveness, effectiveness, and efficiency. The purpose of this stage is to measure user response to artificial intelligence assessment, especially in improving objectivity and consistency in the assessment process, thus ensuring that student evaluation is based on established learning standards. The implementation of artificial intelligence assessment not only supports administrative efficiency for teachers but also provides fair assessment for students according to their learning styles. Before the trial, 20 teachers were asked to create an account on the artificial intelligence assessment SMP Lab UM page. Furthermore, teachers were asked to read artificial intelligence assessment user guide on the website.

Based on the data shown in Figure 9, the artificial intelligence assessment that was piloted showed an average eligibility rate of 94.4%. This figure indicates that the artificial intelligence assessment at the usage trial stage has met the "very feasible" criteria. Responses from teachers indicate that the artificial intelligence assessment is very helpful in evaluating student learning. Teachers feel that the application of

artificial intelligence assessment has become easier. This artificial intelligence assessment is very feasible for implementation in the learning evaluation process by teachers in schools.

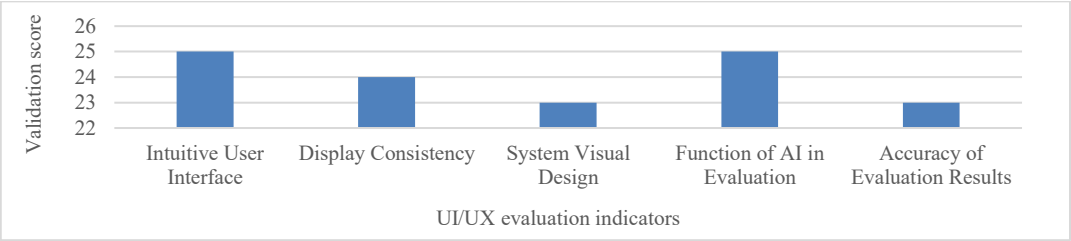

Figure 7. UI/UX validation results

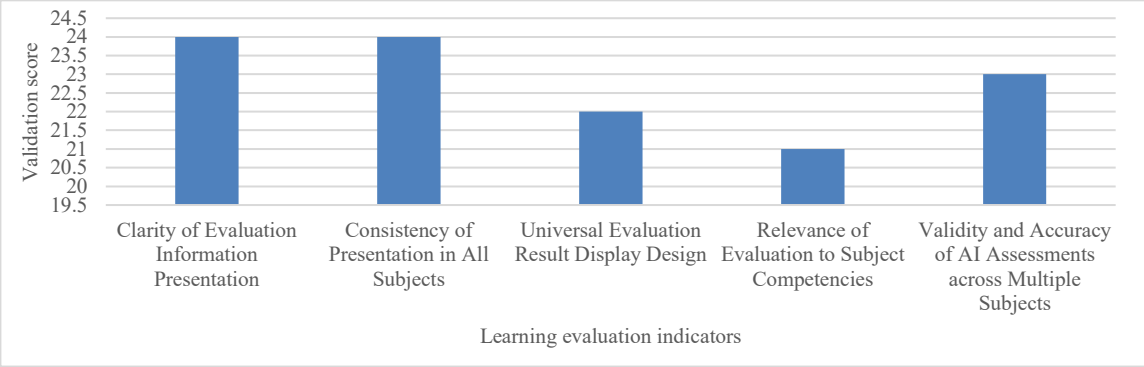

Figure 8. Learning evaluation expert validation results

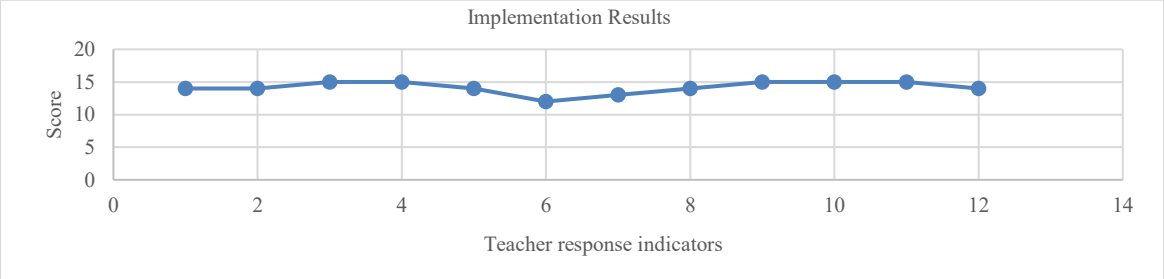

Figure 9. Graph of teacher responses to artificial intelligence assessment

3.5. Evaluation

This evaluation is carried out to ensure that every step of the artificial intelligence assessment development process goes according to plan. In the analysis stage, after analyzing the complexity and needs of teachers, the results obtained are evaluated to ensure that the artificial intelligence assessment developed is relevant and meets the needs of teachers in schools. Furthermore, the evaluation at the design stage is carried out through a review of the formulation of the development objectives and storyboard for the artificial intelligence assessment, with the aim of ensuring that all design elements support the effectiveness of learning evaluation. In the development stage, expert validators assessed the suitability of the artificial intelligence assessment using questionnaires and identified weaknesses such as UI/UX errors, website incompatibilities, and the lack of artificial intelligence's ability to provide high-level thinking analysis in learning evaluation. In the implementation stage, the questionnaire was filled out by teachers who had used the artificial intelligence assessment. Teachers stated that artificial intelligence assessment helped them in assessing students who had different learning characteristics so that it was fair to other students, but there were still shortcomings, such as the absence of a direct score on how much to give to students, but it was still an analysis so that teachers had to score from the results of the artificial intelligence analysis. This evaluation was carried out at each stage of

development to ensure that the artificial intelligence assessment developed was in accordance with the ADDIE model and could produce an effective evaluation tool that met the needs of teachers in schools.

The results of the artificial intelligence assessment website development show that this system still needs further development to cover various subjects. Many important features have not been implemented, which are essential to improving functionality and usability of system in a wider educational environment. For example, the list of questions feature that allows teachers to create, manage, and save questions for various subjects has not been optimized. Swiecki *et al.* [22] mentions that an effective artificial intelligence-based assessment system must provide various types of questions so that it can be adapted to various subjects and learning styles of students. The history chatbot and artificial intelligence assessment features have not been fully developed. This feature is very important because it allows recording of interactions between teachers and students through chatbots, which can be further analyzed to understand student learning patterns and the responses provided by the system. Recording interactions with artificial intelligence allows teachers to more easily identify learners' difficulties and provide timely intervention [23], [24].

Another feature that needs to be developed is the correction of learners' answers in various file formats such as doc, pdf, images, audio, and video. The ability to correct these various formats is very important in the digital era, where learners' tasks are increasingly diverse. The flexibility to accept and correct answers in various formats will increase the effectiveness and accuracy of the artificial intelligence system in assessing learners' work more relevantly [25]. Through a literature review of artificial intelligence-based information systems, several key features that need to be implemented have been identified. These features are further explained in Figure 10, which illustrates the details of features that are still in the development stage and those that have been implemented.

Based on Figure 10, there are still many features that have not been fully implemented. For example, the learning list feature is still limited to general subjects, whereas if it were developed to cover all subjects, such as mathematics, it would simplify the assessment process for teachers from various disciplines, as stated by Ding *et al.* [26], that the application of artificial intelligence in various subjects increases assessment efficiency. In addition, the learning question list feature can only assess essay questions and does not yet support project assessments such as infographics, videos, or other multimedia assignments. According to Newhouse [27], important in providing a comprehensive assessment of technology-based tasks. The history feature of artificial intelligence assessment and chatbot also does not function optimally because after the user logs out, the assessment results are not saved automatically, even though saving previous assessment data is very important for tracking student progress. For example, social studies teachers can assess history reports in pdf form, handmade map images, or video presentations, which support more comprehensive and authentic assessments of student abilities.

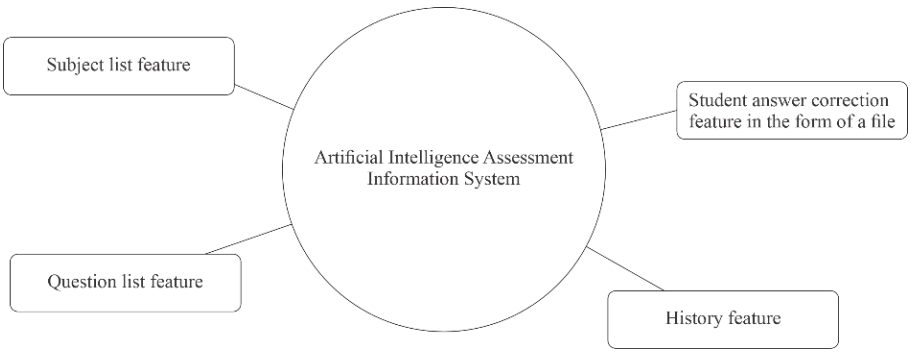

Figure 10. Mind map artificial intelligence assessment information system features that need to be developed

4. CONCLUSION

This study contributes to the field of educational technology by developing a valid and practical artificial intelligence-based learning evaluation system using the ADDIE model. The system, which integrates a virtual assistant chatbot and automated grading artificial intelligence assessment, was proven effective with a feasibility score of 94.4% in field trials involving secondary school teachers. The primary contribution lies in its ability to reduce teachers' administrative burden and provide objective, personalized feedback based on student learning styles. However, this study has several limitations. First, the current question database and learning list are predominantly focused on social studies. Second, the artificial intelligence assessment capability is currently limited in processing complex multimedia formats such as audio and video and maintaining long-term session history. Future work should focus on expanding the system's database to cover a wider range of subjects beyond social studies. Additionally, further development

is needed to enhance the artificial intelligence algorithm to interpret diverse file formats and to improve the history feature for better longitudinal tracking of student progress.

ACKNOWLEDGMENTS

We would like to express our gratitude to the Institute for Research and Community Service at Universitas Negeri Malang for the support of research grant funds which enabled this article to be completed on time.

FUNDING INFORMATION

This article was funded by the Internal of State University of Malang through the 2024 Research Grant Program based on UM Chancellor Decree number 3.4.93/UN32KP/2024.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Adip Wahyudi	✓	✓		✓		✓	✓		✓			✓		✓
Slamet Arifin	✓			✓	✓					✓	✓		✓	
Mukhammad Solikhin	✓		✓		✓			✓		✓	✓			✓
Alfi Sahrina		✓		✓		✓	✓	✓		✓	✓			✓
Feri Fahrian Maulana	✓		✓			✓	✓		✓			✓	✓	
Mohammad Firzon		✓	✓		✓		✓	✓	✓		✓		✓	
Ainur Roziqin														
Maria Emerita		✓	✓			✓		✓		✓		✓		✓
Indraningrum Shrestha														

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

ETHICAL APPROVAL

The research related to human use has been complied with all the relevant national regulations and institutional policies in accordance with the tenets of the Helsinki Declaration and has been approved by the authors' institutional review board or equivalent committee.

DATA AVAILABILITY

The data that support the findings of this study are openly available in https://drive.google.com/drive/folders/1iaXTwYcFY2r945WiGWpocNu9v89GGz9A?usp=drive_link.

REFERENCES

[1] C. Wang *et al.*, "Application of artificial intelligence for feature engineering in education sector and learning science," *Alexandria Engineering Journal*, vol. 110, pp. 108–115, Jan. 2025, doi: 10.1016/j.aej.2024.09.100.
[2] S. Joksimovic, D. Ifenthaler, R. Marrone, M. D. Laat, and G. Siemens, "Opportunities of artificial intelligence for supporting complex problem-solving: Findings from a scoping review," *Computers and Education: Artificial Intelligence*, vol. 4, 2023, doi: 10.1016/j.caeai.2023.100138.
[3] C. Scarpazza and A. Zangrossi, "Artificial intelligence in insanity evaluation. Potential opportunities and current challenges," *International Journal of Law and Psychiatry*, vol. 100, May 2025, doi: 10.1016/j.ijlp.2025.102082.
[4] A. Wahyudi *et al.*, "Integration of micro e-learning content in SIPEJAR geoedupreneurship course based on real life entrepreneurship," *Jurnal Penelitian dan Pengembangan Pendidikan*, vol. 9, no. 1, pp. 56–66, Apr. 2025, doi: 10.23887/jppp.v9i1.86241.

[5] I. Lifshits and D. Rosenberg, "Artificial intelligence in nursing education: a scoping review," *Nurse Education in Practice*, vol. 80, Oct. 2024, doi: 10.1016/j.nepr.2024.104148.
[6] R. Mohan, G. H. S. Kidiya, R. A. Daniel, S. K. Rithana, and A. S. Arjun, "Adaptive intelligence: reshaping education through AI in Ed-Tech," in *Technological Innovation and Enhancing Decision Processes in Modern Enterprises*, Cham, Switzerland: Springer, 2026, pp. 625–638, doi: 10.1007/978-3-032-04378-8_56.
[7] T. Netland, O. V. Dzengelevski, K. Tesch, and D. Kwasnitschka, "Comparing human-made and AI-generated teaching videos: an experimental study on learning effects," *Computers & Education*, vol. 224, Jan. 2025, doi: 10.1016/j.compedu.2024.105164.
[8] N. Spatola, "The efficiency-accountability tradeoff in AI integration: effects on human performance and over-reliance," *Computers in Human Behavior: Artificial Humans*, vol. 2, no. 2, Aug. 2024, doi: 10.1016/j.chbah.2024.100099.
[9] H. Khandakar, S. A. Fazal, K. F. Afnan, and K. K. Hasan, "Implications of artificial intelligence chatbot models in higher education," *IAES International Journal of Artificial Intelligence*, vol. 13, no. 4, pp. 3808–3813, Dec. 2024, doi: 10.11591/ijai.v13.i4.pp3808-3813.
[10] Y. Zheng, H. Meng, and W. Jia, "Application research and challenges of artificial intelligence in primary and secondary education," in *2022 12th International Conference on Information Technology in Medicine and Education*, Nov. 2022, pp. 162–166, doi: 10.1109/ITME56794.2022.00043.
[11] I. Saputra, A. Kurniawan, M. Yanita, E. Y. Putri, and M. Mahniza, "The evolution of educational assessment: how artificial intelligence is shaping the trends and future of learning evaluation," *The Indonesian Journal of Computer Science*, vol. 13, no. 6, Dec. 2024, doi: 10.33022/ijcs.v13i6.4465.
[12] Y. K. Dwivedi *et al.*, "Opinion paper: 'So what if ChatGPT wrote it?' multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy," *International Journal of Information Management*, vol. 71, Aug. 2023, doi: 10.1016/j.ijinfomgt.2023.102642.
[13] V. J. Owan, K. B. Abang, D. O. Idika, E. O. Etta, and B. A. Bassey, "Exploring the potential of artificial intelligence tools in educational measurement and assessment," *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 19, no. 8, Aug. 2023, doi: 10.29333/ejmste/13428.
[14] C. G. Demartini, L. Sciascia, A. Bosso, and F. Manuri, "Artificial intelligence bringing improvements to adaptive learning in education: a case study," *Sustainability*, vol. 16, no. 3, Feb. 2024, doi: 10.3390/su16031347.
[15] I. H. Sarker, "Machine learning: algorithms, real-world applications and research directions," *SN Computer Science*, vol. 2, no. 3, May 2021, doi: 10.1007/s42979-021-00592-x.
[16] G. D. Sahakian, C. Buléon, and G. Alinier, "Educational foundations of instructional design applied to simulation-based education," in *Clinical Simulation: Education, Operations, and Engineering*, Academic Press, 2019, pp. 185–206, doi: 10.1016/B978-0-12-815657-5.00014-0.
[17] A. D. L. R. Gómez, J. M. M. Cano, and G. A. M. Díaz, "Validation of a rubric to evaluate open educational resources for learning," *Behavioral Sciences*, vol. 9, no. 12, Nov. 2019, doi: 10.3390/bs9120126.
[18] S. Kusal, S. Patil, J. Choudrie, K. Kotecha, S. Mishra, and A. Abraham, "AI-based conversational agents: a scoping review from technologies to future directions," *IEEE Access*, vol. 10, pp. 92337–92356, 2022, doi: 10.1109/ACCESS.2022.3201144.
[19] T. P. Li *et al.*, "Socratic artificial intelligence learning (SAIL): the role of a virtual voice assistant in learning orthopedic knowledge," *Journal of Surgical Education*, vol. 81, no. 11, pp. 1655–1666, Nov. 2024, doi: 10.1016/j.jsurg.2024.08.006.
[20] M. E. S. Simaremare, C. Pardede, I. N. I. Tampubolon, D. A. Simangunsong, and P. E. Manurung, "The penetration of generative AI in higher education: a survey," in *2024 IEEE Integrated STEM Education Conference (ISEC)*, Mar. 2024, pp. 1–5, doi: 10.1109/ISEC61299.2024.10664825.
[21] X. Ye and Q. Wang, "Can teacher-student relationship improve the cognitive ability of left-behind children in China? the mediating role of teacher support and learning attitude," *Learning and Instruction*, vol. 93, Oct. 2024, doi: 10.1016/j.learninstruc.2024.101975.
[22] Z. Swiecki *et al.*, "Assessment in the age of artificial intelligence," *Computers and Education: Artificial Intelligence*, vol. 3, 2022, doi: 10.1016/j.caeai.2022.100075.
[23] M. F. Shahzad, S. Xu, W. Naveed, S. Nusrat, and I. Zahid, "Investigating the impact of artificial intelligence on human resource functions in the health sector of China: a mediated moderation model," *Heliyon*, vol. 9, no. 11, Nov. 2023, doi: 10.1016/j.heliyon.2023.e21818.
[24] F. Karataş, B. Eriçok, and L. Tanrikulu, "Reshaping curriculum adaptation in the age of artificial intelligence: mapping teachers' AI-driven curriculum adaptation patterns," *British Educational Research Journal*, vol. 51, no. 1, pp. 154–180, Feb. 2025, doi: 10.1002/berj.4068.
[25] R. Gruetzemacher and J. Whittlestone, "The transformative potential of artificial intelligence," *Futures*, vol. 135, Jan. 2022, doi: 10.1016/j.futures.2021.102884.
[26] A.-C. E. Ding, L. Shi, H. Yang, and I. Choi, "Enhancing teacher AI literacy and integration through different types of cases in teacher professional development," *Computers and Education Open*, vol. 6, Jun. 2024, doi: 10.1016/j.caeo.2024.100178.
[27] C. P. Newhouse, "Digital forms of assessment in schools: supporting the processes to improve outcomes," in *Learning, Design, and Technology*, Cham, Switzerland: Springer, 2023, pp. 1–29, doi: 10.1007/978-3-319-17461-7_41.

BIOGRAPHIES OF AUTHORS

Adip Wahyudi is a geography lecturer at Department of Geography, Faculty of Social Sciences, Universitas Negeri Malang, specializing in Social Geography and Geography Education, who earned his Dr. from Universitas Gadjah Mada in 2020. In addition to his active teaching duties, he is dedicated to developing teaching materials and conducting research, particularly regarding the use of digital technology in geography education to prepare students to face the challenges of the digital age. He can be contacted at email: adip.wahyudi.fis@um.ac.id.

Slamet Arifin is a lecturer in Elementary Education at Teacher Professional Education Program, Graduate School, Universitas Negeri Malang who teaches science and mathematics education. He earned his Ph.D. from the Universitas Pendidikan Indonesia in 2021. His expertise is primarily focused on curriculum development, as well as teaching and assessment methods for science and mathematics at the elementary through secondary levels. In addition to teaching, this lecturer, holding the rank of Penata III/c, is actively engaged in educational research based at the UM Graduate School. He can be contacted at email: mrmasarifin@gmail.com or slamet.arifin.pasca@um.ac.id.

Mukhammad Solikhin is a lecturer in the Department of Mathematics at the Faculty of Mathematics and Natural Sciences, Universitas Negeri Malang, who earned his master's degree (M.Sc.) from the Bandung Institute of Technology (ITB) in 2015. His expertise covers applied mathematics, algebra, and logic and foundations of mathematics, with a deep research interest in the application of mathematics and logic across various scientific fields as well as the development of their underlying theories. He can be contacted at email: mukhammad.solikhin.fmipa@um.ac.id.

Alfi Sahrina is a lecturer in the Department of Geography Education at the Faculty of Education and Social Sciences, Universitas Negeri Malang, who earned her master's degree (M.Ed.) from the same institution in 2015. Holding the rank of Penata III/c, she specializes in innovation in geography education, with a research focus on curriculum development and the implementation of technology to enhance the quality of adaptive learning. She can be contacted at email: alfisahrina13@gmail.com or alfi.sahrina.fis@um.ac.id.

Feri Fahrian Maulana is an academic who holds a master's degree in Environmental Science from Environmental Resource Management and Development Program, Graduate School, Universitas Brawijaya and specializes in technology-based education and environmental development. His research focuses on spatial analysis, mapping, and field research to understand the impacts of environmental and economic changes on communities. He can be contacted at email: ferifahrian@student.ub.ac.id.

Mohammad Firzon Ainur Roziqin is an assistant professor at the Department of Electrical Engineering and Informatics, Faculty of Engineering, Universitas Negeri Malang, specializing in the development of learning technologies that integrate artificial intelligence, augmented reality (AR), and virtual reality (VR). He is dedicated to introducing interactive and practical teaching innovations to equip students with critical thinking skills and competencies relevant to the industry in the digital age. He can be contacted at email: firzonainur@gmail.com.

Maria Emerita Indraningrum Shrestha has served as associate dean at The British College and vice principal of Department of Business Administration, Kathmandu College of Management, Nepal. She is strongly committed to academic excellence and innovative teaching, focusing on developing a deep understanding of management and social sciences among students. His expertise spans across various aspects of management, including business management, marketing, and human resources, enabling her to provide a comprehensive and interdisciplinary education. She is dedicated to enhancing critical thinking, analytical skills, and practical understanding of students, and preparing them for the challenges of a dynamic business world. She can be contacted at email: mariashrestha@yahoo.com or m.shrestha@thebritishcollege.edu.np.