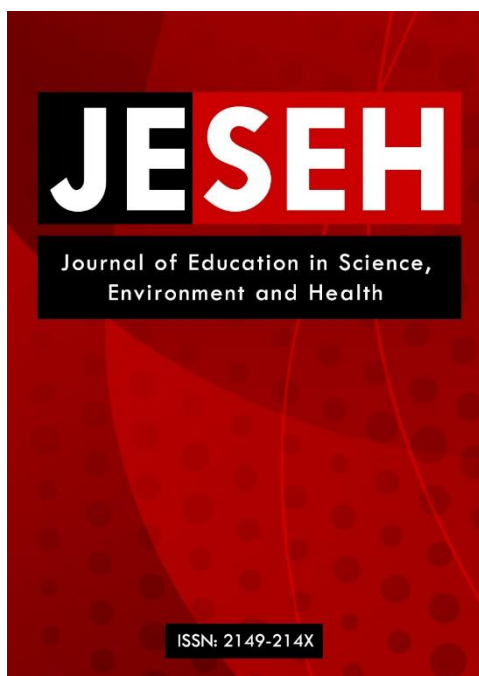




ISSN: 2149-214X

**Journal of Education in Science,
Environment and Health**

www.jeseh.net

Use of Artificial Intelligence in Biology Education: A Systematic Review of Literature

Tugce Duran, Musa Dikmenli
Necmettin Erbakan University

To cite this article:

Duran, T., & Dikmenli, M. (2025). Use of artificial intelligence in biology education: a systematic review of literature. *Journal of Education in Science, Environment and Health (JESEH)*, 11(4), 314-333. <https://doi.org/10.55549/jeseh.864>

This article may be used for research, teaching, and private study purposes.

Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

Authors alone are responsible for the contents of their articles. The journal owns the copyright of the articles.

The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of the research material.

Use of Artificial Intelligence in Biology Education: A Systematic Review of Literature

Tugce Duran, Musa Dikmenli

Article Info

Article History

Published:
01 October 2025

Received:
18 May 2025

Accepted:
12 September 2025

Keywords

Biology education
Artificial intelligence
Systematic review

Abstract

In recent years, the use of artificial intelligence (AI) has become increasingly widespread and has attracted significant attention worldwide. In this study, a systematic review was conducted to determine the use of AI in biology education and the prevailing trends in its application. The study aimed to conduct a comprehensive review of articles published between 2010 and 2024 that employed artificial intelligence in biology education. In line with this aim, a total of 49 keywords were searched from Web of Science, SCOPUS, ERIC and IEEE Xplore databases. The entire process was summarized in the PRISMA diagram. 39 articles were deemed eligible for inclusion in the systematic review. The selected articles were analyzed in terms of publication year, research method, country of conduct, study group, number of participants, subject area, artificial intelligence technologies used, artificial intelligence applications used, and their outcomes. An evaluation of the articles according to the main topics in biology revealed that there is a lack of sufficient research on the evolutionary history of biodiversity and the models and processes of evolution. It was also found that there is a lack of sufficient research in the literature on artificial intelligence-supported educational games and simulations in biology education. In this context, it is recommended that the use of artificial intelligence technologies in biology education be expanded to include educational games, which are frequently used to motivate students and encourage learning, and simulations, which are suitable for development in many areas of biology education.

Introduction

First, in 1950, Alan M. Turing proposed considering the question “*Can a machine think?*” in his article titled “Computing Machinery and Intelligence.” In this context, he presented a thought experiment called the “Turing test” to bring together the concepts of thinking and machines in order to demonstrate that a machine can think or, in other words, exhibit human-like intelligence (Turing, 1950). In 1959, Prof. Dr. Cahit Arf, in his article “*Can a Machine Think and How Can It Think?*”, presented and explained examples of machine design demonstrating that machines can think. According to Arf, machines can be designed to perform analytical and logical operations such as establishing analogies, using language, calculating, and eliminating. Therefore, there are similarities between the human brain and machine functioning. However, Arf argued that the most fundamental differences between the human brain and a machine stem from the human brain's ability to function with aesthetic awareness, to make decisions, and to feel free to choose whether or not to carry out a given task (Arf, 1959). It is reported in the literature that the difficulty in determining the parameters of artificiality or in identifying the reasons why machines differ from human intelligence makes it difficult to define artificial intelligence, and the following statement is made regarding this issue;

“They are much less than human intelligence—they can only calculate. And they are much more—they can calculate larger numbers and faster than humans. We have cause to be in awe at the super-human brilliance of their feats of calculation.” (Cope et al., 2021).

As can be understood, machines are emphasized as possessing superhuman intelligence in calculations. Artificial intelligence can be defined as the ability of machines to exhibit and simulate human-like intelligent behavior. In other words, it can be defined as software used to perform tasks or produce outputs that are considered to require human intelligence (Oxford University Press, n.d.).

There are several significant milestones in the historical development of artificial intelligence. The first of these was Marvin Minsky (1969) and John McCarthy (1971), who laid the foundations of the field based on representation and reasoning. McCarthy, the founder of the term artificial intelligence, received the Turing Award

for his contributions. Later, Allen et al. (1975) established the foundations of artificial intelligence with their study on symbolic models of human cognition and problem solving; Ed Feigenbaum and Raj Reddy (1994) pioneered the development of expert systems that aim to solve real-world problems by encoding human knowledge; Judea Pearl (2011) developed probabilistic reasoning techniques and integrated them into artificial intelligence; and finally, Yoshua Bengio, Geoffrey Hinton, and Yann LeCun (2019) made deep learning an essential part of modern technology. These prominent figures, who significantly contributed to the development of artificial intelligence, were also awarded the Turing Award (Erman, Hayes-Roth, et al., 1980; Feigenbaum, & McCorduck, 1983; LeCun et al., 2015; Luckin, et al., 2016; McCarthy, 1987; Minsky & Papert, 1969; Newell, & Simon, 1976; Pearl, 2009).

Artificial intelligence can be described as an umbrella term encompassing numerous technologies and applications. Language models, one of the AI technologies, facilitate tasks such as grammar assistance, question answering, search engine response optimization, text generation, and translation. However, it is often difficult and complex to distinguish the texts generated by language models from those produced by humans. This can lead to academic fraud, deliberate misuse, and plagiarism. Therefore, while language models offer significant benefits, they also present challenges (Brown, et al., 2020). If AI is used effectively, all the challenges brought about by AI technologies can be overcome with the power of human intelligence (Akintande, 2024).

Cognitive tutors, one of the AI applications, support students' learning processes by providing personalized feedback and contribute to long-term learning. In this regard, cognitive tutors can be considered a powerful educational tool. However, it should be noted that AI can never replace real teachers, as the functioning and operations of AI are fundamentally different from human intelligence (Koedinger & Corbett, 2005). Machines cannot substitute for teachers, but they can serve as supportive tools (Crovello, 1974).

In their study, Lu, et al. (2024) found that chatbots (ChatGPT), an application of artificial intelligence, can be used to score students' short-answer questions and demonstrate good-to-moderate consistency when compared to teacher scorings. Similarly, Jukiewicz (2024) used ChatGPT to evaluate student assignments and found positive correlation between these AI-based evaluations and teacher evaluations. The study concluded that ChatGPT can be used as an effective tool for grading student assignments, considering its high-quality assessment, unbiased grading, time-saving, and feedback-generating capabilities. Elgohary and Al-Dossary (2022) determined that the use of artificial intelligence-supported virtual classrooms significantly improved the field training and teaching skills of female teacher candidates (84.40%). Almeda, et al. (2018) developed artificial intelligence-supported models that predict students' course success on an online learning platform. The study found that these models performed quite well in predicting student success. Predicting student success is crucial for providing support to students identified as being at-risk. Accordingly, Mubarak et al. (2022) developed a machine learning-based prediction model for early identification of students at risk of dropping out. As a result of the study, the use of this model enabled the identification of at-risk students with an accuracy rate of 84%. Benhamdi et al. (2017) presented a recommendation approach that provides personalized learning materials for e-learning environments based on students' preferences, memory capacities, interests, and readiness. They found that this recommendation approach increases the quality of learning. Ijaz et al. (2017) combined artificial intelligence and virtual reality to create a virtual replica of the city of Uruk and used AI-controlled 3D avatars to recreate daily life. They found that this application, which allowed students to walk the streets of this city and talk to its residents, resulted in increased motivation and interest in their learning experiences. Aluthman (2016) examined the effects of the AI technology-based Criterion® system, which employs natural language processing (NLP), on the writing performance of students enrolled in an academic writing course in the English Language Department at a university. This system, which provides instant feedback, evaluation, and automatic scoring, was found to improve students' writing mechanics, with moderate progress in style, grammar, and usage. Koć-Januchta et al. (2020) developed a digital biology textbook using AI-supported question-and-answer technologies and visuals. The study revealed that students' engagement in asking questions and interacting with visuals was positively correlated with retention. The usability of this digital textbook was perceived positively by students. The use of artificial intelligence in education is becoming increasingly widespread (Holmes et al., 2023). In the field of education, artificial intelligence can measure knowledge, support learning, and enable automatic transfer between numbers and meaning. In this context, AI holds promise for the future in education and assessment. However, educators should be aware of the inherent limitations of AI (Cope et al., 2021). It is evident that AI has a significant impact on teaching and learning both within the educational sector and in educational institutions (Chen et al., 2020).

Artificial intelligence is used in a wide range of fields, including industry, marketing, financial services, engineering, medicine, pharmacy, physical education, physics education, chemistry education, science education, biology education, mathematics education, and language teaching (Broussard et al., 2019; Cooper 2023; Ding et al., 2023; Fernández, 2019; Hamet & Tremblay, 2017; Hessler & Baringhaus, 2018; Holmes et al., 2004; Iyamuremye et al., 2024; Jarek & Mazurek, 2019; Miller et al., 2025; Nasution, 2023; Parunak, 1996; Pham &

Pham, 1999; Xu et al., 2022). Artificial intelligence technologies such as deep learning are used to examine and categorize biological data (Webb, 2018). In general, artificial intelligence in biology is used in areas including disease detection and diagnosis, medication management, personalized medicine, biological data analysis, synthetic biology, investigating and integrating complex mechanisms at various scales, bioinformatics, radiography, image processing, and genetic data analysis (Aripin et al., 2024; Bhardwaj et al., 2022; Hassoun et al., 2021). The use of artificial intelligence is considered to potentially cause a revolutionary change in biology in the 21st century (Hassoun, 2021).

The use of computers in biology education helps improve teaching, makes it possible to teach difficult topics, increases students' interest in the course, reduces tedious tasks related to simple topics, and allows students to learn at their own pace and review course materials as often as they wish. In this context, the use of computers in biology education can improve teaching quality. However, excessive use should be avoided, and optimization should always be ensured in computer use (Crovello, 1974).

There are numerous systematic reviews on the use of artificial intelligence: AI in education (Wang et al., 2024; Zhai et al., 2021), AI in student assessment (González-Calatayud et al., 2021), AI and learning analytics in teacher education (Salas-Pilco et al., 2022), AI technologies in K-12 education (Martin, Zhuang, & Schaefer, 2024), the use of ChatGPT in K-12 education (Zhang & Tur, 2024), AI applications in online higher education (Ouyang, Zheng, & Jiao, 2022), AI in English language teaching (Sharadgah & Sa'di, 2022), AI in science education (Almasri, 2024), AI in science teaching and learning (Heeg & Avraamidou, 2023), AI in biology and biology learning (Aripin et al., 2024), and bibliometric analyses on the quality and role of AI in improving biology education (Lidiastuti et al., 2025). However, studies specifically focusing on the use of AI in biology education are relatively limited. Therefore, compiling and presenting the literature on the use of artificial intelligence in biology education, which has become increasingly widespread in recent years and has made a significant impact worldwide, is considered important in determining the status and trends in the use of AI in biology education. This study aimed to conduct a systematic review by comprehensively examining articles published between 2010 and 2024 to determine the current status and trends in the use of artificial intelligence in biology education. Accordingly, the present study is expected to provide a general overview of AI use in biology education and contribute to the existing literature.

Purpose of the Study and Sub-Problems

The purpose of this study is to determine how artificial intelligence is used in biology education and to identify trends related to its use. Accordingly, answers were sought to the following sub-questions:

- 1) Which artificial intelligence technologies are used in studies on the use of artificial intelligence in biology education?
- 2) Which artificial intelligence applications are used in studies on the use of artificial intelligence in biology education?
- 3) What are the outcomes of studies on the use of artificial intelligence in biology education?
- 4) What is the distribution of studies on the use of artificial intelligence in biology education by year?
- 5) What is the distribution of studies on the use of artificial intelligence in biology education by research method?
- 6) What is the distribution of studies on the use of artificial intelligence in biology education by country?
- 7) What is the distribution of studies on the use of artificial intelligence in biology education by study group and the number of participants?
- 8) What is the distribution of studies on the use of artificial intelligence in biology education by the number of participants?
- 9) What subject areas do studies on the use of artificial intelligence in biology education focus on?

Method

A systematic review is a method that allows for the comprehensive and systematic screening of published studies in a given field, using various inclusion and exclusion criteria to answer research questions and problems. What distinguishes systematic reviews from other types of literature reviews is that they are comprehensive, objective, and reproducible. Their reproducibility stems from the fact that the researcher explicitly specifies the search terms, databases, and the inclusion and exclusion criteria at the beginning of the study. This also indicates that the systematic reviews are evidence-based. Systematic reviews are therefore regarded as important studies that minimize bias and yield reliable findings (Higgins & Green, 2008; Karaçam, 2013; Page et al., 2021; Zawacki-

Richter, 2020). In this study, a systematic review was conducted to determine the use of artificial intelligence technologies, which are becoming increasingly widespread in education, in biology education and to identify the current trends in this field. A five-phase systematic review process was followed to address the research problems:

Phase 1: Article Collection, Review, and Initial Selection

Databases and Search Terms, Article Collection

To review the relevant literature, four international databases (Web of Science, SCOPUS, ERIC, and IEEE Xplore) were searched for articles. For each database, the terms "artificial intelligence" and "biology education" were searched in the entire text (all fields). These terms were searched by combining them using AND or +. To access all the data, the search strings were expanded. Seven different alternative terms for "artificial intelligence" and seven different alternative terms for "biology education" were added. By crossing these strings with each other, a total of 49 searches were conducted in each database. All search strings used are presented in Table 1. The database search and downloading of relevant studies were completed between May and June 2025.

Table 1. Search strings used to search databases

Topic	Search string
Artificial intelligence	"artificial intelligence" OR "machine learning" OR "AI" OR "natural language processing" OR "deep learning" OR "artificial neural networks" OR "expert systems"
AND	"biology education" OR "biology learning" OR "biology teaching" OR
Biology education	"biology instruction" OR "biology curriculum" OR "biology laboratory" OR "biology textbook"

Article Review and Initial Selection

All articles retrieved after searching the databases were uploaded to Zotero. A separate collection was created for each database in Zotero. All collections were then compiled into a single collection under the name "Combined Folder." The articles in this collection were reviewed, and duplicate articles were excluded. The remaining articles were then evaluated for eligibility according to the predefined inclusion criteria. The articles were first reviewed by their titles, then by their abstracts, and finally by their full texts, independently by two authors. Disagreements between the two authors were resolved through discussion.

Initial Inclusion Criteria

Six criteria were applied to determine the eligibility of studies for inclusion in this study: (1) Being appropriate for biology education content; (2) Not being a book, book chapter, conference proceeding, or thesis; (3) Being empirical research; (4) Being written in English; (5) Having been conducted between 2010 and 2024. Therefore, articles that were not published between 2010 and 2024, were not empirical, were not written in English, and were not appropriate for biology education were not included in this study. In addition, books, book chapters, conference proceedings, and these were not included in this study.

The inclusion and exclusion procedures employed in this systematic review were summarized using the PRISMA diagram (Moher et al., 2009) (Figure 1). The articles retrieved from Web of Science (n=220), SCOPUS (n=9,121), ERIC (n=68), and IEEE Xplore (n=77) databases were combined into one folder. 3,335 duplicate articles encountered in different databases were excluded, leaving 6,151 articles. First, the titles of these articles were screened, and 4,934 articles deemed outside the scope of the study were excluded. Then, the abstracts of the remaining articles were screened, and 1,002 articles were excluded for being irrelevant to the scope of the study. Finally, the full texts of the remaining 215 articles were examined in detail and evaluated according to the initially determined eligibility criteria. As a result of the evaluation, 176 articles were excluded based on the eligibility criteria: (1) 102 articles identified as being from fields such as physical education, medicine, nursing, pharmacy, chemistry, and physics; (2) 19 articles identified as being written for purposes such as systematic review, meta-analysis, compilation, and program promotion, and therefore not empirical; (3) 6 articles identified as conference proceedings; and (4) 49 articles identified as being published between 1989 and 2025 were excluded. Since all reviewed articles were written in English, no exclusion was made based on language criteria.

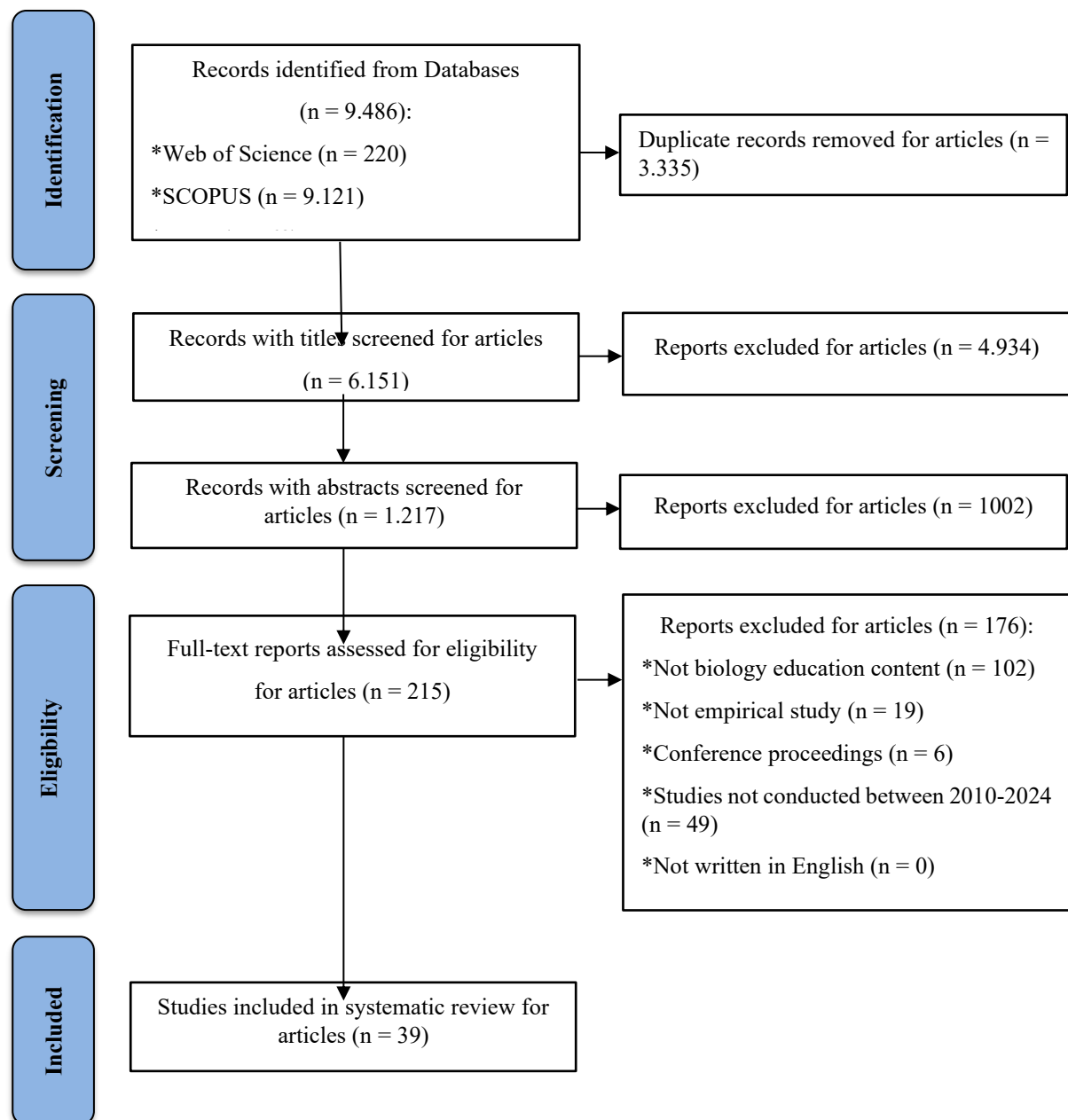


Figure 1. PRISMA diagram

As shown in the PRISMA diagram in Figure 1, a total of 9,486 articles were initially retrieved from the databases. After excluding duplicate articles encountered in different databases, the titles of the remaining 6,151 articles were independently screened by two authors. A 94% (5,790 / 5,790 + 361) agreement was reached between the authors regarding the articles to be included in the study. After the necessary eliminations, the abstracts of the remaining 1,217 articles were independently screened by the two authors, resulting in a 97% (1,185 / 1,185 + 32) agreement regarding the articles to be included in the study. Finally, the full texts of the remaining 215 articles were independently screened by two authors, resulting in a 99% (213 / 213 + 2) agreement regarding the articles to be included in the study based on the eligibility criteria. In order to ensure the reliability of the study, these agreement rates were calculated according to the formula of Miles and Huberman (2016).

Phase 2: Final Article Selection

Artificial intelligence is a broad field encompassing a wide range of technologies, including machine learning, natural language processing, computer vision, generative AI, expert systems, robotic systems, deep learning, large language models, and natural language generation. Each technology is further divided into sub-applications. In

this study, 39 articles selected for full-text review were examined in detail and classified according to their AI technologies.

Phase 3: Data Determination

For the 39 articles included in the study, a table was created in Excel to determine the following characteristics: (1) publication year, (2) research method, (3) country of conduct, (4) study group, (5) number of participants, (6) subject area, (7) AI technologies used, (8) AI applications used, and (9) outcomes. The authors independently listed the characteristics to be examined in the articles. Any disagreements between the authors were then reviewed, and the lists were revised accordingly. Ultimately, agreement was reached between the authors regarding the dataset to be used in the study.

Phase 4: Data Extraction and Audit

Following the selection of articles to be included in the study and the determination of data, all excluded articles were removed from Zotero through the joint effort of the two authors. Additionally, the dataset was reviewed by a professor specializing in the field of biology education to ensure data accuracy. Finally, the data were verified, and the final dataset was prepared.

Phase 5: Analysis

This study aimed to address nine sub-problems. Descriptive analysis was used to analyze articles on the use of artificial intelligence in biology education based on the AI technologies used, publication year, research method, country of conduct, study group, and number of participants. In descriptive analysis, the dataset is categorized according to pre-determined themes. Descriptive analysis is carried out in four stages: (1) creating a framework for descriptive analysis, (2) processing the data according to the thematic framework, (3) defining the findings, (4) interpreting the findings (Yıldırım & Şimşek, 2016, pp. 239-240).

Content analysis was used to analyze articles on the use of artificial intelligence in biology education according to the artificial intelligence applications used, the outcomes, and the subject area. Content analysis is carried out in four stages: (1) coding the data, (2) identifying themes, (3) organizing codes and themes, and (4) defining and interpreting the findings (Yıldırım & Şimşek, 2016, pp. 242-252). The data set used in this study was coded by generating codes directly from the data using inductive analysis in accordance with the “coding based on concepts extracted from the data” type (Strauss & Corbin, 1990). Excel, IBM-SPSS 24, and MAXQDA 2018 programs were used in the analysis and presentation of the data.

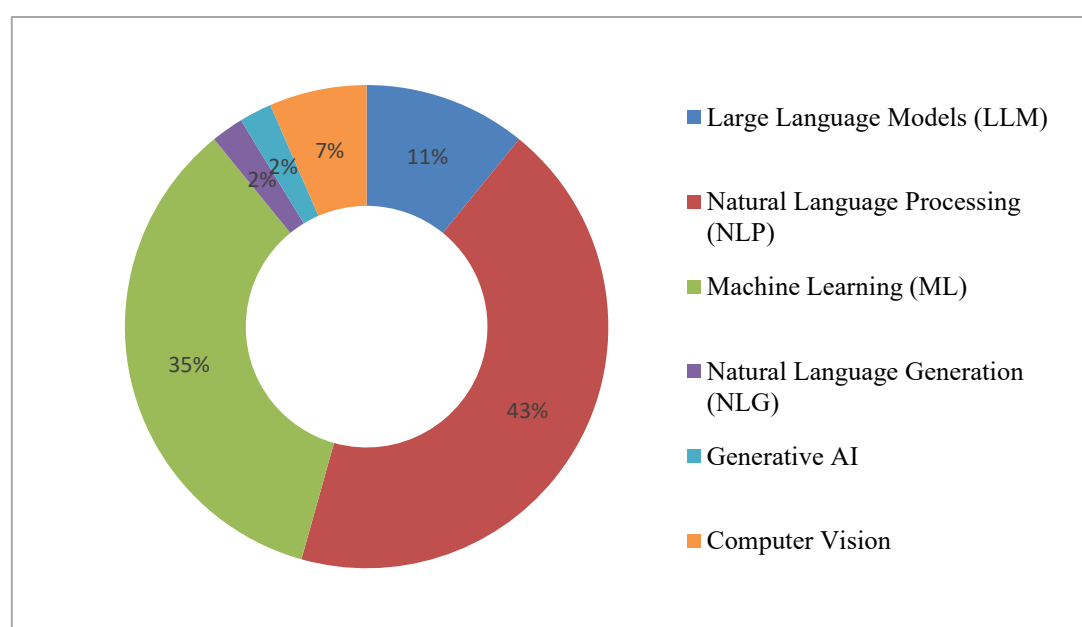


Figure 2. Artificial intelligence technologies used in the reviewed articles

Results

Descriptive analysis was employed to analyze articles on the use of artificial intelligence in biology education according to the AI technologies used. Some articles employed more than one AI technology (Ariely et al., 2023; Chaudhri et al., 2013; Ha et al., 2011; Jho & Ha, 2024; Royse et al., 2024; Sripathi et al., 2023; Zhang & VanLehn, 2016). Therefore, all AI technologies were evaluated separately. As a result, the total number obtained represented the total number of AI technologies used ($f = 46$). The distribution of AI technologies used according to the analysis results is shown in the doughnut chart (Figure 2). As shown in Figure 2, "Natural Language Processing (NLP)" ($f=20$; 43%) was the most frequently used AI technology in articles using AI in biology education, followed by "Machine Learning (ML)" ($f=16$; 35%) and "Large Language Models (LLM)" ($f=5$; 11%). The least frequently used AI technologies in these studies were "Natural Language Generation (NLG)" ($f=1$; 2%) and "Generative AI" ($f=1$; 2%).

In this study, content analysis was used to analyze articles on the use of artificial intelligence in biology education according to the AI application used. The AI applications used in 39 articles were listed. In some articles, more than one AI application was employed (Aleksandrovich et al., 2024; Ceylan & Karakuş, 2024; Chaudhri et al., 2013; Chen & Liu, 2024; Coglianò et al., 2022; Koć-Januchta et al., 2020; Koć-Januchta et al., 2022; Peffer et al., 2020; Zafeiropoulos & Kalles, 2024; Zhang & VanLehn, 2016). Therefore, all AI applications were evaluated separately, and the total number obtained represented the total number of AI applications used. The AI applications used in the reviewed articles were coded independently by the two authors. A total of 56 codes were generated. At this stage, 95% ($53 / 53 + 3$) agreement was reached between the authors (Miles & Huberman, 2016).

Table 2. Artificial intelligence applications used in the reviewed articles

Themes	Codes	<i>f</i>	%
Chatbots and Question-Answer Systems	Educational Chatbots ($f=11$)	16	28,5 7
	Question-Answer Technology ($f=2$)		
	Educational Question-Answer Systems ($f=1$)		
	Web-Based Question Compilation ($f=1$)		
	Knowledge-Based Question Generation ($f=1$)		
	Automated Assessment Systems ($f=9$)		
Automated Assessment and Feedback	Automated Computer-Scoring Model ACSM) ($f=1$)	13	23,2 1
	Constructed-Response Classifier-CRC ($f=1$)		
	Summarization Integrated Development Environment (SIDE) ($f=1$)		
	An Online Formative Assessment Tool Called "Evograder" ($f=1$)		
	Neural Networks ($f=2$)		
Basic Artificial Intelligence Techniques and Algorithms	Data Clustering and Network Analysis ($f=2$)	9	16,0 7
	Genetic Algorithms ($f=1$)		
	Text Mining ($f=1$)		
	Computerized Lexical Analysis ($f=1$)		
	Text Classification ($f=1$)		
	Bayesian Structure Learning ($f=1$)		
Knowledge-Based Systems	Knowledge Representation ($f=3$)	7	12,5 0
	Knowledge-Acquisition ($f=2$)		
	Knowledge Base ($f=2$)		
Learning Analytics and Predictive Models	Learning Analytics ($f=2$)	4	7,14
	Predictive Learning Analytics ($f=1$)		
	Predictive Modeling ($f=1$)		
Image Processing and Multimodal Interaction	Image Recognition Technologies ($f=3$)	4	7,14
	Multimodal Interaction Design ($f=1$)		
Intelligent Tutoring Systems and Personalized Learning	Virtual Tutors ($f=1$)	2	3,57
	Personalized Assistants ($f=1$)		
Educational Games	Educational Computer Game ($f=1$)	1	1,79
Total		56	100

Similar codes were combined to create themes. The themes were determined as follows: (1) Chatbots and Question-Answer Systems, (2) Automated Assessment and Feedback, (3) Basic Artificial Intelligence Techniques and Algorithms, (4) Knowledge-Based Systems, (5) Learning Analytics and Predictive Models, (6) Image

Processing and Multimodal Interaction, (7) Intelligent Tutoring Systems and Personalized Learning, and (8) Educational Games. Expert opinion was consulted to ensure the accuracy and consistency of the codes with the themes. The expert was provided with two separate lists: one containing the codes and the other containing the themes and was asked to match the codes with the themes. According to the results, the agreement was calculated as 96% (54 / 54 + 2) (Miles & Huberman, 2016). For codes and themes where disagreements occurred, agreement was reached through discussion. All codes, themes, and their frequencies are presented in Table 2.

An examination of Table 2 reveals that in articles on the use of artificial intelligence in biology education, AI applications are clustered under eight themes, each consisting of 28 codes with a total frequency of 56. The theme with the highest frequency was "*Chatbots and Question-Answer Systems*" ($f=16$), followed by "*Automated Assessment and Feedback*" ($f=13$), and "*Basic Artificial Intelligence Techniques and Algorithms*" ($f=9$). The highest-frequency code within the "*Chatbots and Question-Answer Systems*" theme was "Educational chatbots" ($f=11$). Articles using educational chatbots used platforms such as ChatGPT, Bard/Gemini, BingChat/Microsoft Copilot, and YouChat. The code "Automated assessment systems" for the theme "*Automated Assessment and Feedback*" represented articles that did not specifically specify the name of the program used, but simply included the general phrase "Automated assessment systems." If a program, such as EvoGrader, was explicitly mentioned, the program name itself was used as the code. The theme with the lowest frequency was "Educational Games", represented by only one article. This article discussed the use of a machine learning-based educational computer micro-game as a teaching tool (Brom et al., 2011).

Table 3. Outcomes of the reviewed articles

Themes	Codes	f	%
Educational Technology & Tool Development Outcomes	Feedback & Smart Guidance ($f=8$)	20	32,7 9
	Usability ($f=2$)		
	Smart Microscope Design ($f=1$)		
	Smart Textbook Development ($f=1$)		
	Digital Textbook Use ($f=1$)		
	Textbook Analysis ($f=1$)		
	Creating a Virtual Collection ($f=1$)		
	Creating a Virtual Laboratory ($f=1$)		
	Digital Assistance/Guidance ($f=1$)		
	Visual Analysis/Measurement Automation ($f=1$)		
	Supporting Fieldwork ($f=1$)		
	Modeling Learning Progress ($f=1$)		
	Student success ($f=3$)		
Cognitive Learning Outcomes	Knowledge acquisition and retention ($f=3$)	12	19,6 7
	Conceptual understanding and changes ($f=2$)		
	Detection of misconceptions ($f=1$)		
	Learning gain ($f=1$)		
	Students' knowledge retention and transfer ($f=1$)		
	Systems thinking skills ($f=1$)		
	Motivation ($f=5$)		
Affective & Motivational Outcomes	Student perception ($f=2$)	11	18,0 3
	Student engagement and satisfaction ($f=1$)		
	Student attitude ($f=1$)		
	Epistemological beliefs about science-(EBAS) ($f=1$)		
	Feeling and thought analysis ($f=1$)		
Teacher & Institutional Outcomes	Teacher Workload & Assessment Quality ($f=8$)	10	16,3 9
	Effectiveness & Sustainability of the Program/Department ($f=2$)		
	Cognitive load ($f=2$)		
Metacognitive & Strategic Outcomes	Cognitive strategy use ($f=1$)	4	6,56
	Self-regulation ($f=1$)		
	Assessing Question Answering Performance ($f=3$)		
Assessment & Feedback Outcomes	Assessing Question Writing Quality ($f=1$)	4	6,56
Total		61	100

In this study, content analysis was conducted to analyze articles on the use of artificial intelligence in biology education according to their outcomes. The outcomes of these articles were identified and noted. Since some

articles aimed at multiple outcomes (Aleksandrovich et al., 2024; Ceylan & Karakuş, 2024; Chen & Liu, 2024; Jho & Ha, 2024; Kim & Kim, 2022; Koć-Januchta et al., 2022; Koć-Januchta et al., 2020; Uhl et al., 2021; Wang, et al., 2019; Yin et al., 2024; Zafeiropoulos & Kalles, 2024), all outcomes were evaluated separately. Therefore, the total number reached at the end of the analysis represents the total number of outcomes, not the total number of articles. All reviewed articles were coded in terms of outcomes by two authors. A total of 61 codes were generated. A 98% (60 / 60 + 1) agreement was reached between the authors (Miles & Huberman, 2016). The generated codes were grouped among themselves to create themes. The themes were (1) Educational Technology and Tool Development Outcomes, (2) Cognitive Learning Outcomes, (3) Affective and Motivational Outcomes, (4) Teacher and Institutional Outcomes, (5) Metacognitive and Strategic Outcomes, and (6) Assessment and Feedback Outcomes. To validate the code–theme compatibility, expert opinions were obtained from two professors working at the faculty of education in a state university. Each expert was provided with separate lists of codes and themes and asked to match the codes with the themes so that none remained unmatched. Accordingly, the experts' agreement was determined as 97% (59 / 59 + 2) and 98% (60 / 60 + 1), respectively. In the final analysis, agreement was reached through discussion for the codes and themes where disagreements occurred. All codes, themes, and their frequencies obtained from the analysis are presented in Table 3.

An examination of Table 3 revealed that the outcomes of articles on the use of artificial intelligence in biology education were grouped under 6 themes, consisting of 32 codes with a total frequency of 61. The theme with the highest frequency was “*Educational Technology and Tool Development Outcomes*” ($f=20$), followed by “*Cognitive Learning Outcomes*” ($f=12$) and “*Affective and Motivational Outcomes*” ($f=11$). The themes with the lowest frequency were “*Metacognitive and Strategic Outcomes*” ($f=4$) and “*Assessment and Feedback Outcomes*” ($f=4$). The code with the highest frequency within the theme “*Educational Technology and Tool Development Outcomes*” was “*Feedback & Intelligent Guidance*” ($f=8$). This code represented articles that used platforms/applications that provided instant personalized feedback to students, and articles that guided students in using an AI-supported smart microscope that included a physical interaction kit (Ariely et al., 2023, 2024; Jho & Ha, 2024; Wang et al., 2019; Yin et al., 2024; Zafeiropoulos & Kalles, 2024). The most frequently observed outcomes in the “*Cognitive Learning Outcomes*” theme were “*Student success*” ($f=3$) and “*Knowledge acquisition and retention*” ($f=3$), while the most common outcome in the “*Affective and Motivational Outcomes*” theme was “*Motivation*” ($f=5$). In the theme “*Teacher and Institutional Outcomes*”, the code with the highest frequency was “*Teacher Workload & Assessment Quality*” ($f=8$), which represented articles that specifically aimed to reduce teacher workload in assessing written responses from large student populations by using AI-based tools (Beggrow et al., 2014; Beigman et al., 2017; Ha et al., 2011; Haudek et al., 2012; Jescovitch et al., 2021; Jho & Ha 2024; Moharreri et al., 2014; Sripathi et al., 2023). Finally, within the “*Assessment and Feedback Outcomes*” theme, the code “*Evaluating Question-Answering Performance*” ($f=3$) represented articles evaluating the scientific question answering performance of chatbots such as ChatGPT (Dao & Le, 2023; Elmas et al., 2024; Crowther et al., 2023). A descriptive analysis was conducted to determine the distribution of the 39 articles reviewed in this study by publication year. Based on the results of this analysis, a column chart was created to show the distribution of the articles by publication year (Figure 3).

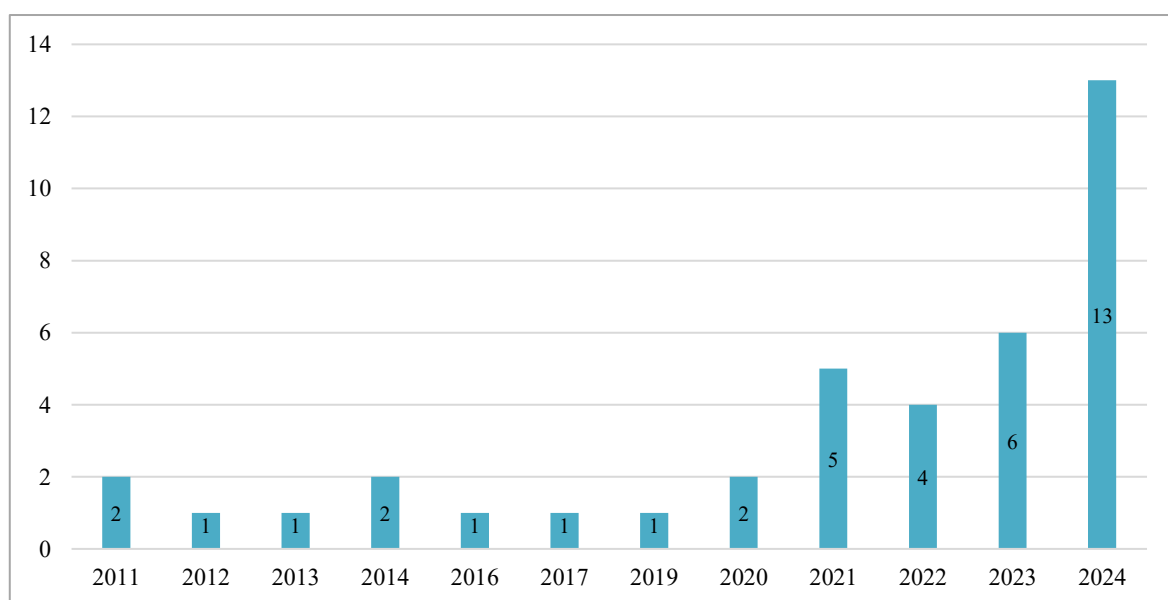


Figure 3. Distribution of reviewed articles by publication year

As shown in Figure 3, the number of articles on the use of artificial intelligence in biology education has steadily increased over the years, with a sharp rise observed in 2021. In particular, 2024 witnessed more than a twofold increase compared to the previous year. Therefore, 2024 was the year with the highest number of studies ($f=13$; 33.33%). However, no studies were found in 2015 and 2018. A descriptive analysis was conducted to determine the distribution of reviewed articles by research methods. Based on the analysis results, a pie chart was created to show the distribution of articles by research method (Figure 4).

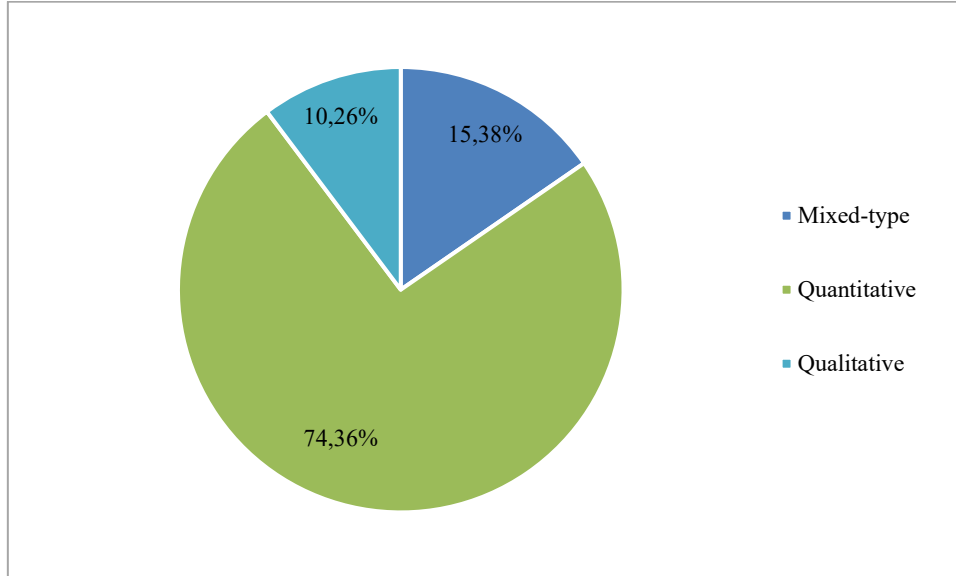
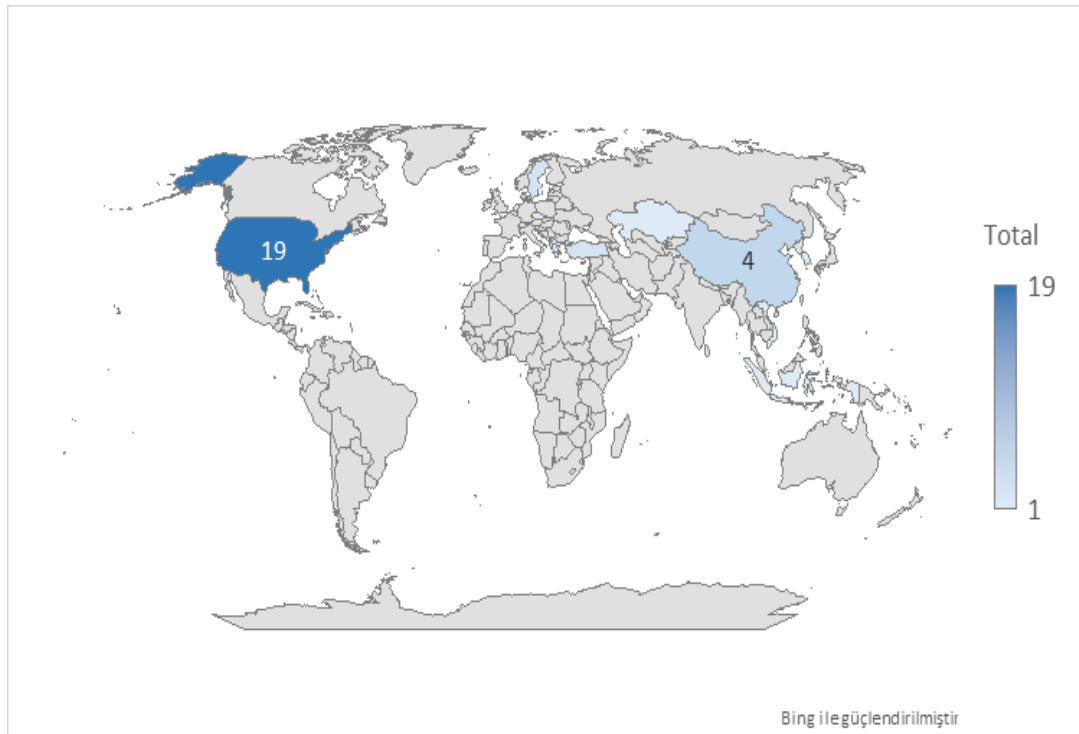


Figure 4. Distribution of reviewed articles by research method



USA	19	Indonesia	1
China	4	South Korea	1
Taiwan	3	England	1
Israel	2	Kazakhstan	1
Sweden	2	Vietnam	1
Türkiye	2	Greece	1
Czech Republic	1		

Figure 5. Distribution of reviewed articles by country

Figure 4 shows that quantitative methods were generally used in articles on the use of artificial intelligence in biology education ($f=29$; 74.36%). Mixed research methods were the second most preferred method ($f=6$; 15.38%). The least commonly used research method in these studies was qualitative research methods ($f=4$; 10.26%). A descriptive analysis was conducted to determine the distribution of the reviewed articles by country. Based on the analysis, the countries where the articles were conducted and their frequencies were visualized on a world map. On the map, the number of articles conducted in each country was highlighted using different shades of color, and the frequencies of articles conducted in each country were also presented descriptively (Figure 5). Figure 5 shows that articles on the use of artificial intelligence in biology education were conducted in 13 countries (USA, China, Taiwan, Israel, Sweden, Turkey, Czech Republic, Indonesia, South Korea, UK, Kazakhstan, Vietnam, and Greece). The country with the most articles was the US (48.72%), followed by China (10.26%) and Taiwan (7.69%).

Descriptive statistics were used to determine the distribution of the articles by study group. Brock et al. (2024) analyzed textbooks in their study, and therefore, the study group in their article was analyzed as “textbooks.” Dao and Le (2023) evaluated the performance of various large language models in answering biology exam questions, and the study group in their article was analyzed as “large language model applications (ChatGPT, Microsoft Bing Chat, Google Bard).” Crowther et al. (2023) examined the performance differences of chatbots based on large language models, and the study group in their article was analyzed as “chatbot versions (ChatGPT, Google Bard, YouChat).” Elmas et al. (2024) evaluated the validity of the responses produced by ChatGPT when asked scientific questions, and the study group in their article was analyzed as “ChatGPT.” Some articles were found to have been conducted on more than one study group (Dao & Le, 2023; Peffer, et al., 2020; Wang et al., 2019; Zafeiropoulos & Kalles, 2024). Therefore, all study groups were evaluated separately. Consequently, the total number obtained in the analysis represented the total number of study groups. The results of the analysis are presented in Table 4.

Table 4. Distribution of reviewed articles by study group

Study Group	<i>f</i>	%
University students	22	47,83
High school students	6	13,04
Middle school students	4	8,70
ChatGPT	3	6,52
Google Bard	2	4,35
Faculty members	2	4,35
Biology graduates	1	2,17
Teachers	1	2,17
Textbooks	1	2,17
YouChat	1	2,17
Biology experts	1	2,17
Post-secondary students	1	2,17
Microsoft Bing Chat	1	2,17
Total	46	100

An examination of Table 4 revealed that the study group of articles on the use of artificial intelligence in biology education was composed primarily of university students ($f=22$; 47.83%), followed by high school ($f=6$; 13.04%) and middle school students ($f=4$; 8.70%). Artificial intelligence applications such as ChatGPT, YouChat were also considered as study groups and had a significant proportion ($f=7$; 15.21%). The distribution of articles by the number of participants was determined using descriptive statistics through the IBM-SPSS 24 program. In the study conducted by Brock et al. (2024), the number of textbooks reviewed was considered as the number of participants. Some articles evaluated the question performance of artificial intelligence technologies, so the number of questions was considered as the number of participants (Dao & Le, 2023; Elmas, Adiguzel-Ulutas et al., 2012). The results of the analysis are presented in Table 5.

Table 5. Data on the number of participants in the reviewed articles

	N	Min.	Max.	Mean	Std. Deviation
Number of participants	39	5	4937	498,62	1013,78

Table 5 shows that the number of participants in articles on the use of artificial intelligence in biology education ranged from a minimum of 5 to a maximum of 4,937. On average, articles on the use of artificial intelligence in biology education had 499 participants.

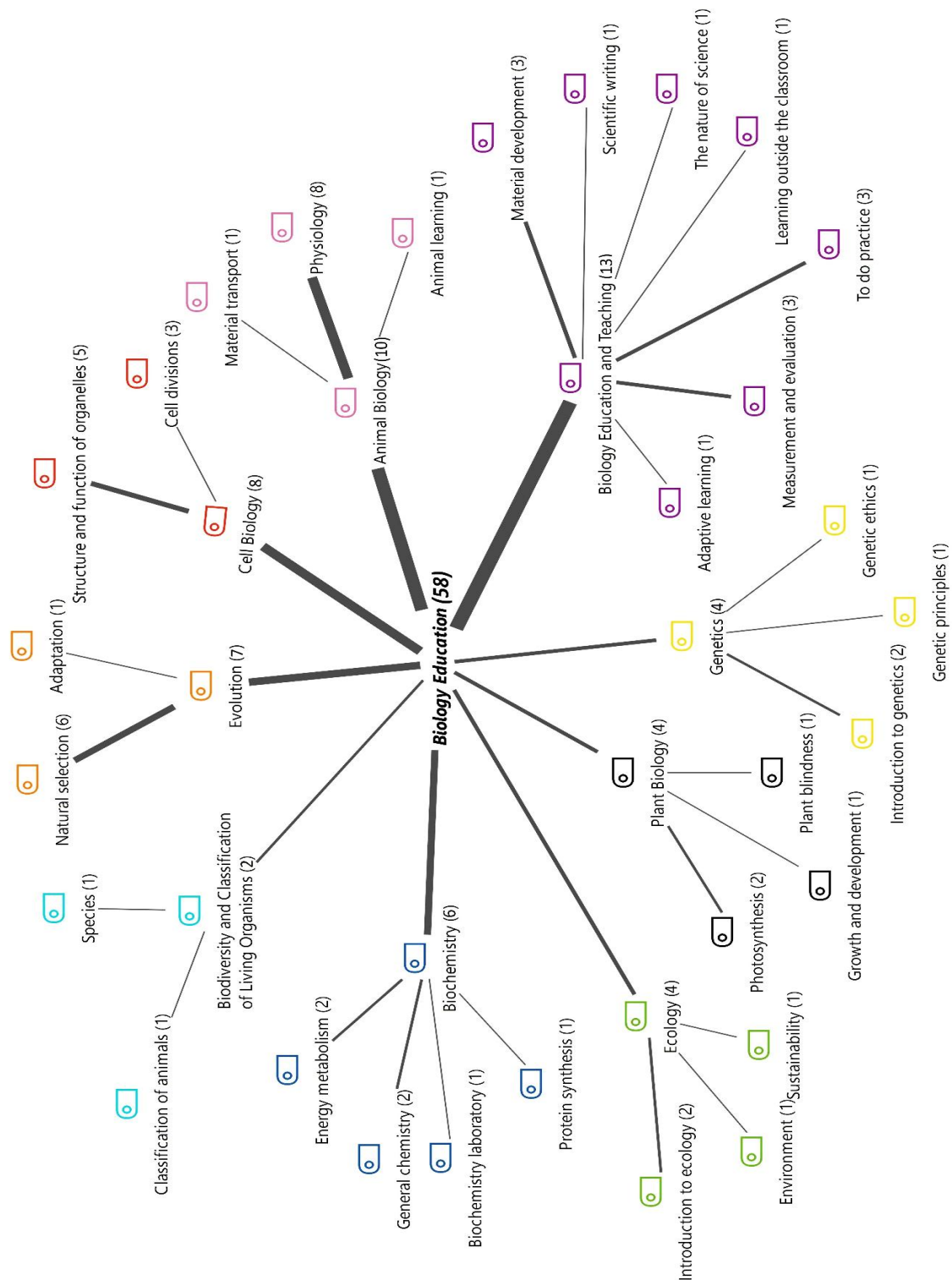


Figure 6. Distribution of reviewed articles by subject area

Content analysis was used to analyze articles using AI in biology education according to their subject areas. In an Excel table, the subject areas of the 39 reviewed studies were listed. Some articles focused on more than one subject area (Ariely et al., 2024; Beigman et al., 2017; Brock et al.2024; Ceylan & Karakuş, 2024; Chaudhri et al., 2013; Crowther et al., 2023; Dao & Le, 2023; Elmas et al., 2024; Koć-Januchta et al., 2020; Koć-Januchta et al., 2022; Lin & Ye, 2023). Therefore, all subject areas were evaluated separately. Consequently, the total number obtained in the analysis represented the total number of subject areas, not the total number of articles. The subject areas in the analyzed articles were coded independently by both authors. A total of 58 codes were generated after negotiation. A 98% (57 / 57 + 1) agreement was reached between the authors (Miles & Huberman, 2016). Similar codes were grouped to form themes. The themes were organized according to basic topics in biology education.

The themes were composed of concepts that best represented the codes. The themes were determined as (1) Biology Education and Teaching, (2) Animal Biology, (3) Cell Biology, (4) Evolution, (5) Biochemistry, (6) Genetics, (7) Plant Biology, (8) Ecology, and (9) Biodiversity and Classification of Living Organisms. Two experts, both professors in the biology education department at a state university, were consulted to ensure that the codes accurately represented these themes. The professors were given lists of codes (listed alphabetically) and themes (with brief descriptions). They were asked to independently match the codes to the themes. The Miles and Huberman (2016) formula was used to determine reliability. Accordingly, the agreement of the experts was determined as 98% ($57 / 57 + 1$) and 95% ($55 / 55 + 3$), respectively. Agreement was reached on the theme and code matching where there was disagreement.

The 58 codes and 9 themes identified as a result of the analysis were transferred to the MAXODA 2018 program. After completing the necessary coding, a map was created using MAXMaps. This map was based on the "Code-Subcode-Departments model." In this model, the themes represented the codes, the codes represented the subcodes, and biology education represented the department. All analyzed articles were related to biology education. Therefore, inclusiveness was taken into account in the selection of the department name. In the created map, each theme and its related codes were shown in a different color. The line widths of all connections in the map reflected the frequencies (Figure 6).

Figure 6 shows that studies on the use of artificial intelligence in biology education were grouped under 9 themes based on subject area, and these themes consisted of 29 codes with a total frequency of 58. The theme with the highest frequency was "Biology Education and Teaching" ($f=13$), followed by "Animal Biology" ($f=10$) and "Cell Biology" ($f=8$). In the "Biology Education and Teaching" theme, the codes with the highest frequencies were "To do practice," "Measurement and evaluation," and "Material development." In the "Animal Biology" theme, the code with the highest frequency was "Physiology." In the "Cell Biology" theme, the code with the highest frequency was "Structure and function of organelles." The theme with the lowest frequency was "Biodiversity and Classification of Living Organisms," which included two codes with the equal frequencies: "Classification of animals" and "Species."

Conclusions and Recommendations

In this study, a systematic review was conducted to determine the use of artificial intelligence in biology education and related trends. A total of 49 keywords were searched from the Web of Science, SCOPUS, ERIC, and IEEE Xplore databases. All articles retrieved from the search were stored in Zotero. Inclusion and exclusion procedures were applied based on the initially established criteria, first by title, then by abstract, and finally by full text. The entire process was summarized in the PRISMA diagram. 39 articles were included in the systematic review. The included articles were analyzed in terms of publication year, research method, country of conduct, study group, number of participants, subject area, artificial intelligence technologies used, artificial intelligence applications used, and outcomes.

Natural Language Processing (NLP) was found to be the most frequently used AI technology in articles on the use of artificial intelligence in biology education, followed by Machine Learning (ML) and Large Language Models (LLM). In their study, Salas-Pilco et al. (2022) examined the studies conducted between 2017 and 2021 on the use of artificial intelligence and learning analytics in teacher education. They reported that ML was the most commonly used artificial intelligence technology in the articles they reviewed. The results of this study are similar to the results of our study. In a study in which a systematic literature review was conducted on the use of artificial intelligence in English language teaching, articles published between 2015 and 2021 were analyzed. The analysis revealed that the AI technologies used in the reviewed articles were NLP, data mining, deep learning, decision tree, ML, cloud computing and edge computing, support vector machine, expert system, neural network, and genetic algorithms (Sharadgah & Sa'di, 2022). The results of this study are similar to the results of the current study in terms of the use of NLP and ML. Ouyang et al. (2022) conducted a study aiming to provide an overview of AI applications in online higher education. Designed as a systematic literature review, this study included studies using artificial intelligence in online higher education between 2011 and 2020. The analysis of the selected articles revealed that Decision Tree, Neural Network, Naive Bayes, and Support Vector Machine were the most frequently used AI Technologies in these articles. The usage rate of NLP technologies was determined to be 6.25%. The results of this study contradict the results of the present study. The different context and application areas focused on in the study by Ouyang et al. (2022) are considered to have a decisive impact on the types and usage rates of the AI technologies used. In studies using AI technologies in biology education, NLG and Generative AI are among the least preferred AI technologies. However, the use of Generative AI in educational contexts can offer many advantages. Specifically, it allows for the creation of personalized learning systems

customized to students' learning styles and individual needs (Holmes et al., 2023). Similarly, GenAI models can be effectively used to produce interactive educational materials, enrich learning experiences, and simulate educational scenarios (Sengar et al., 2025). Therefore, it was observed that Generative AI technologies are used in only a limited number of studies in biology education, indicating a need for further research in this area.

The analysis of articles on the use of artificial intelligence in biology education according to the AI application used revealed that the theme with the highest frequency was "Chatbots and Question-Answer Systems," followed by "Automated Assessment and Feedback" and "Basic Artificial Intelligence Techniques and Algorithms" within the "Chatbots and Question-Answer Systems" theme, the code with the highest-frequency was "Educational chatbots." Articles using educational chatbots used platforms such as ChatGPT, Bard/Gemini, BingChat/Microsoft Copilot, and YouChat. Research on the use of ChatGPT in teaching and learning indicates that it offers numerous advantages, including advanced communication capabilities, versatility, natural language processing, performance evaluation, and text generation enhancement. However, the use of ChatGPT in teaching and learning also has several disadvantages, including error detection issues, plagiarism and originality concerns, privacy and data security risks, dependency, response quality, and bias (Ali et al., 2024). Another study investigating the use of ChatGPT in K-12 education similarly emphasized that ChatGPT offers significant advantages, such as facilitating educators' roles and responsibilities, creating instructional materials, lesson planning, and optimizing student learning experiences through personalized learning, but also drawbacks related to ethics, data privacy, and academic dishonesty. Additionally, the use of ChatGPT in K-12 education is considered potentially promising (Zhang & Tur, 2024). ChatGPT is seen as an effective AI tool for designing units, assessment criteria and exams in the field of science (Cooper, 2023). In a study analyzing articles using AI technologies in K-12 education between 2017 and 2022, it was reported that the AI technology applications used in the articles included virtual reality devices, machine learning modeling tools, chatbots, AI robots, and smart teachers (Martin et al., 2024). The results of this study are similar to the results of the current study. In a systematic review of studies on AI use in science education between 2014 and 2023, Almasri (2024) found that AI was used in areas such as exam creation, assessment, improving the learning environment, and predicting academic performance.

The results of the study conducted by Almasri (2024) are consistent with the themes identified in the current study (*Chatbots and Question-Answer Systems, Automated Assessment and Feedback, Learning Analytics and Predictive Models*). Aripin et al. (2024), in their study on the use of artificial intelligence in biology and biology learning, identified AI technology models used in biology education as adaptive modeling, experience point data modeling, interactive books, smart classrooms, and virtual laboratories. In this context, they indicated that AI in biology learning encompasses assessment and evaluation, instructional media, virtual classrooms, enrichment of learning, teaching assistance, and learning aids. These categories determined by Aripin et al. (2024) are consistent with the themes identified in the current study (*Educational Games, Intelligent Tutorial Systems and Personalized Learning, Knowledge-Based Systems, Automated Assessment and Feedback*). Similarly, in a systematic review of studies on AI use in science teaching and learning between 2010 and 2021, it was found that the most frequently used AI applications were automated assessment and feedback, predictive modeling, and personalized learning (Heeg & Avraamidou, 2023). These categories align with the themes identified in the present study (*Automated Assessment and Feedback, Learning Analytics and Predictive Models, and Intelligent Tutorial Systems and Personalized Learning*).

Based on the analysis of articles on the use of artificial intelligence in biology education within the scope of this study, the theme with the highest frequency was identified as "Educational Technology and Tool Development Outcomes," followed by "Cognitive Learning Outcomes" and "Affective and Motivational Outcomes". The themes with the lowest frequency were "Metacognitive and Strategic Outcomes" and "Assessment and Feedback Outcomes". The most frequent code under the "Teacher and Institutional Outcomes" theme was "Teacher Workload & Assessment Quality," which included AI technologies used to reduce teacher workload in tasks such as reviewing students' written responses (Beggrow et al., 2014; Beigman, et al., 2017; Ha et al., 2011; Haudek et al., 2012; Jescovitch et al., 2021; Jho & Ha, 2024; Moharreri et al., 2014; Sripathi et al., 2023). Teachers can increase efficiency and effectiveness in tasks such as grading student assignments and providing feedback through the use of AI, which in turn leads to improved teaching quality (Chen et al., 2020). In their systematic review of articles on AI use in science teaching and learning, Heeg and Avraamidou (2023) stated that AI applications can alleviate the workload of science educators, increase students' interest in science through personalized learning environments, and optimize teaching processes to improve low learning outcomes in science classes. The categories identified by Heeg and Avraamidou (2023) are consistent with the themes identified in the current study (*Teacher and Institutional Outcomes, Affective and Motivational Outcomes, Cognitive Learning Outcomes, Educational Technology and Tool Development Outcomes*).

The number of articles on the use of artificial intelligence in biology education has gradually increased over the years, with a particularly sudden rise in 2021, and the highest number of articles was conducted in 2024. Lidiastuti et al. (2025) analyzed studies published between 2000 and 2025 through a bibliometric analysis in order to investigate the role of artificial intelligence in improving biology education. According to the results of the study, the use of artificial intelligence in biology education shows an increasing trend over the years, with the most significant increase occurring since 2018. They found that the highest number of articles was conducted in 2023. In a study analyzing articles using artificial intelligence technologies in K-12 education between 2017 and 2022, it was determined that the use of artificial intelligence technologies in K-12 education increased after 2019, with the peak in 2021 (Martin et al., 2024). A systematic review of articles using AI in student assessment from 2010 to 2020 indicated that the number of articles was higher between 2015 and 2020 (González-Calatayud et al., 2021). In a study conducted by Zhai et al. (2021), articles on the use of artificial intelligence in education between 2010 and 2020 were examined and it was determined that the use of artificial intelligence has increased over the years, with the highest number of studies being conducted especially in 2020. Almasri (2024), in a systematic review of AI use in science education from 2014 to 2023, similarly found that AI applications in science education increased over time, with the peak in 2023. Therefore, the results of the current study were found to be consistent with the literature.

In this study, it was determined that quantitative methods were generally used in the articles on the use of artificial intelligence in biology education, while qualitative methods were the least used research methods. Similarly, Zhai et al. (2021), in their study aiming to examine how artificial intelligence is applied in education and the trends in this area, analyzed studies using artificial intelligence in education between 2010 and 2020 and reported that quantitative research was predominant in studies on the use of artificial intelligence in education. In a study analyzing articles on the use of artificial intelligence technologies in K-12 education between 2017 and 2022, it was determined that qualitative methods were generally used in the articles, followed by quantitative methods (Martin et al., 2024). The results of the study by Martin et al. (2024) contradict the results of the current study. This is thought to be due to the difference in their focus areas. While the current study focused on biology education studies, Martin et al. (2024) focused on K-12 education.

In a study examining the articles on the use of artificial intelligence and learning analytics in teacher education between in the current study, it was determined that the most articles on the use of artificial intelligence in biology education were conducted in the United States, followed by China and Taiwan. Lidiastuti et al. (2025), in their bibliometric analysis of AI applications in biology education between 2000 and 2025, similarly reported that the highest number of publications was conducted in the United States, followed by China and Germany. In a systematic review of articles on the use of AI for student assessment between 2010 and 2020, it was determined that most studies were conducted in the United States according to the origins of the article authors (González-Calatayud et al., 2021). In a study analyzing articles on the use of AI technologies in K-12 education between 2017 and 2022, it was reported that most studies were conducted in the United States, followed by Korea and Brazil (Martin et al., 2024). 2017 and 2021, it was found that most studies were conducted in China, followed by the United States, Germany, and Canada (Salas-Pilco et al., 2022). In a study examining the articles on the use of artificial intelligence in science, it was determined that the country where the most studies were conducted was the United States, followed by Germany (Almasri, 2024). Therefore, the results of the current study were found to be consistent with the literature.

It was found that the study group of articles on the use of artificial intelligence in biology education mostly consisted of university students, followed by high school and middle school students. In a systematic review examining the articles on the use of artificial intelligence in education between 2010 and 2020, it was found that the study group in the articles mostly consisted of university students (Zhai et al., 2021). In another systematic review examining AI applications in education from 1984 to 2022, it was found that the study group of nearly half of the articles consisted of higher education students (Wang et al., 2024). In a study examining the articles using artificial intelligence in English language teaching between 2015 and 2021, it was found that the study group in the articles were generally higher education students (Sharadgah & Sa'di, 2022). In a study analyzing the articles on the use of artificial intelligence technologies in K-12 education between 2017 and 2022, it was determined that the study group of the articles consisted mostly of high school students (Martin et al., 2024). In a study examining articles on the use of artificial intelligence and learning analytics in teacher education between 2017 and 2021, it was determined that the study group of the articles generally consisted of pre-service teachers (PSTs) (Salas-Pilco et al., 2022). Almasri (2024), in a systematic review of AI use in science education between 2014 and 2023, found that studies were mostly conducted with undergraduate students, followed by high school and middle school students. Therefore, the results of the present study are consistent with the existing literature.

The number of participants in articles on the use of artificial intelligence in biology education was determined to be minimum 5 and maximum 4,937. The average number of participants in articles on the use of artificial intelligence in biology education was found to be 499. The studies on the use of artificial intelligence in biology education were categorized into 9 themes according to their subject areas [(1) Biology Education and Teaching, (2) Animal Biology, (3) Cell Biology, (4) Evolution, (5) Biochemistry, (6) Genetics, (7) Plant Biology, (8) Ecology and (9) Biodiversity and Classification of Living Organisms]. The theme with the highest frequency was "Biology Education and Teaching", followed by the themes "Animal Biology" and "Cell Biology". When evaluated according to the main topics included in the works "*Campbell Biology*" and "*Life: The Science of Biology*", which are accepted as fundamental in biology education and accepted worldwide, it was seen that there is insufficient research on the evolutionary history of biological diversity and the models and processes of evolution. (Sadava et al., 2014; Urry et al., 2022).

Overall, it can be concluded that the use of artificial intelligence in biology education is becoming increasingly widespread; however, not all technologies and applications are being utilized yet, and studies generally focus on chatbot and response system applications. The literature lacks sufficient studies on AI-supported educational games and simulations in biology education. In this context, it is recommended that the use of AI technologies in biology education be expanded through educational games, which are frequently used to motivate students and encourage learning, and simulations, which can be developed for various topics in biology education. For instance, AI-supported activities can be created to illustrate historical processes and geological periods that people cannot directly experience in their daily lives, such as natural selection, adaptation, and evolution. Additionally, mass extinction events can be simulated using AI. Researchers aiming to conduct studies on the use of AI in biology education are encouraged to address the gaps identified in the literature, specifically focusing on the "evolutionary history of biodiversity" and the "models and processes of evolution."

Scientific Ethics Declaration

* The authors declare that the scientific ethical and legal responsibility of this article published in JESEH journal belongs to the authors.

Conflict of Interest

* The authors declare that they have no conflicts of interest.

Acknowledgements or Notes

* We would like to contribute to the reliability of the study thank the professors specialized.

References

- Akintande, O. J. (2024). Artificial versus natural intelligence: Overcoming students' cheating likelihood with artificial intelligence tools during virtual assessment. *Future in Educational Research*, 2(2), 147-165.
- Aleksandrovich, S. I., Ramazan, T., Utegaliyeva, R., Sarimbayeva, B., Keubassova, G., Bissalyeva, R., Syman, K., & Abdikarimova, G. (2024). Transformative applications in biology education: A case study on the efficacy of adaptive learning with numerical insights. *Caspian Journal of Environmental Sciences*, 22(2), 395-408.
- Ali, D., Fatemi, Y., Boskabadi, E., Nikfar, M., Ugwuoke, J., & Ali, H. (2024). ChatGPT in teaching and learning: A systematic review. *Education Sciences*, 14(6), 643.
- Almasri, F. (2024). Exploring the impact of artificial intelligence in teaching and learning of science: A systematic review of empirical research. *Research in Science Education*, 54(5), 977-997.
- Almeda, M. V., Zuech, J., Utz, C., Higgins, G., Reynolds, R., & Baker, R. S. (2018). Comparing the factors that predict completion and grades among for-credit and Open/MOOC students in online learning. *Online Learning*, 22(1), 1-18.
- Aluthman, E. S. (2016). The effect of using automated essay evaluation on ESL undergraduate students' writing skill. *International Journal of English Linguistics*, 6(5), 54-67.

- Arf, C. (1959). Can a machine think and how can it think?. *Atatürk Üniversitesi 1958-1959 Öğretim Yılı Halk Konferansları*, 1, 91-103.
- Ariely, M., Nazaretsky, T., & Alexandron, G. (2023). Machine learning and hebrew NLP for automated assessment of open-ended questions in biology. *International Journal of Artificial Intelligence in Education*, 33(1), 1-34.
- Ariely, M., Nazaretsky, T., & Alexandron, G. (2024). Causal-mechanical explanations in biology: Applying automated assessment for personalized learning in the science classroom. *Journal of Research in Science Teaching*, 61(8), 1858-1889.
- Aripin, I., Gaffar, A. A., Jabar, M. B. A., & Yulianti, D. (2024). Artificial intelligence in biology and learning biology: A literature review. *Jurnal Mangifera Edu*, 8(2), 41-48.
- Beggrow, E. P., Ha, M., Nehm, R. H., Pearl, D., & Boone, W. J. (2014). Assessing scientific practices using machine-learning methods: How closely do they match clinical interview performance? *Journal of Science Education and Technology*, 23(1), 160-182.
- Behrens, K. A., Marbach-Ad, G., & Kocher, T. D. (2025). AI in the genetics classroom: A useful tool but not a replacement for creative writing. *Journal of Science Education and Technology*, 34, 621-635.
- Beigman - Klebanov, B., Burstein, J., Harackiewicz, J. M., Priniski, S. J., & Mulholland, M. (2017). Reflective writing about the utility value of science as a tool for increasing STEM motivation and retention-Can AI help scale up? *International Journal of Artificial Intelligence in Education*, 27(4), 791-818.
- Ben, S., Liu, C., Yang, P., Gong, J., & He, Y. (2024). A practical evaluation of online self-assisted previewing architecture on rain classroom for biochemistry lab courses. *Frontiers in Education*, 9, 1326284.
- Benhamdi, S., Babouri, A., & Chiky, R. (2017). Personalized recommender system for e-Learning environment. *Education and Information Technologies*, 22(4), 1455-1477.
- Bertolini, R., Finch, S. J., & Nehm, R. H. (2021). Testing the impact of novel assessment sources and machine learning methods on predictive outcome modeling in undergraduate biology. *Journal of Science Education and Technology*, 30(2), 193-209.
- Bhardwaj, A., Kishore, S., & Pandey, D. K. (2022). Artificial intelligence in biological sciences. *Life*, 12(9), 1430.
- Brock, R., Tsourakis, N., & Kampourakis, K. (2024). Using text mining to identify teleological explanations in physics and biology textbooks: An exploratory study. *Science and Education*, 34, 2167-2188.
- Brom, C., Preuss, M., & Klement, D. (2011). Are educational computer micro-games engaging and effective for knowledge acquisition at high-schools? A quasi-experimental study. *Computers and Education*, 57(3), 1971-1988.
- Broussard, M., Diakopoulos, N., Guzman, A. L., Abebe, R., Dupagne, M., & Chuan, C. H. (2019). Artificial intelligence and journalism. *Journalism & Mass Communication Quarterly*, 96(3), 673-695.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877-1901.
- Ceylan, B., & Karakus, M. A. (2024). Development of an artificial intelligence-based mobile application platform: Evaluation of prospective science teachers' project on creating virtual plant collections in terms of plant blindness and knowledge. *International Journal of Technology in Education and Science*, 8(4), 668-688.
- Chaudhri, V. K., Cheng, B. H., Overholtzer, A., Roschelle, J., Spaulding, A., Clark, P., Greaves, M., & Gunning, D. (2013). Inquire biology: A textbook that answers questions. *AI Magazine*, 34(3), 55-72.
- Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access*, 8, 75264-75278.
- Chen, P. Y., & Liu, Y. C. (2024). Impact of ai robot image recognition technology on improving students' conceptual understanding of cell division and science learning motivation. *Journal of Baltic Science Education*, 23(2), 208-220.
- Chuang, C. H., Lo, J. H., & Wu, Y. K. (2023). Integrating chatbot and augmented reality technology into biology learning during COVID-19. *Electronics*, 12(1), 222.
- Cogliano, M., Bernacki, M. L., Hilpert, J. C., & Strong, C. L. (2022). A self-regulated learning analytics prediction-and-intervention design: Detecting and supporting struggling biology students. *Journal of Educational Psychology*, 114(8), 1801-1816.
- Cooper, G. (2023). Examining science education in ChatGPT: An exploratory study of generative artificial intelligence. *Journal of Science Education and Technology*, 32(3), 444-452.
- Cope, B., Kalantzis, M., & Sears-Smith, D. (2021). Artificial intelligence for education: Knowledge and its assessment in AI-enabled learning ecologies. *Educational Philosophy and Theory*, 53(12), 1229-1245.
- Crovello, T. J. (1974). Computers in biological teaching. *BioScience*, 24(1), 20-23.
- Crowther, G. J., Sankar, U., Knight, L. S., Myers, D. L., Patton, K. T., Jenkins, L. D., & Knight, T. A. (2023). Chatbot responses suggest that hypothetical biology questions are harder than realistic ones. *Journal of Microbiology & Biology Education*, 24(3) e00153-23.

- Dao, X. Q., & Le, N. B. (2023). LLMs performance on Vietnamese high school biology examination. *International Journal of Modern Education and Computer Science*, 15(6), 14-30.
- Ding, H., Wu, J., Zhao, W., Matinlinna, J. P., Burrow, M. F., & Tsoi, J. K. (2023). Artificial intelligence in dentistry-A review. *Frontiers in Dental Medicine*, 4, 1085251.
- Elgohary, H. K. A., & Al-Dossary, H. K. (2022). The effectiveness of an educational environment based on artificial intelligence techniques using virtual classrooms on training development. *International Journal of Instruction*, 15(4), 1133-1150.
- Elmas, R., Adiguzel-Ulutas, M., & Yilmaz, M. (2024). Examining ChatGPT's validity as a source for scientific inquiry and its misconceptions regarding cell energy metabolism. *Education and Information Technologies*, 29(18), 25427-25456.
- Erman, L. D., Hayes-Roth, F., Lesser, V. R., & Reddy, D. R. (1980). The Hearsay-II speech-understanding system: Integrating knowledge to resolve uncertainty. *ACM Computing Surveys*, 12(2), 213-253.
- Feigenbaum, E. A., & McCorduck, P. (1983). *The fifth generation: Artificial intelligence and Japan's computer challenge to the world*. Addison-Wesley.
- Fernández, A. (2019). Artificial intelligence in financial services. *Economic Bulletin: Banco de Espana*, 19, 1-7.
- González-Calatayud, V., Prendes-Espinosa, P., & Roig-Vila, R. (2021). Artificial intelligence for student assessment: A systematic review. *Applied Sciences*, 11(12), 5467.
- Ha, M., Nehm, R. H., Urban-Lurain, M., & Merrill, J. E. (2011). Applying computerized-scoring models of written biological explanations across courses and colleges: Prospects and limitations. *CBE-Life Sciences Education*, 10(4), 379-393.
- Hamet, P., & Tremblay, J. (2017). Artificial intelligence in medicine. *Metabolism*, 69, 36 - 40.
- Hassoun, S., Jefferson, F., Shi, X., Stucky, B., Wang, J., & Rosa Jr, E. (2021). Artificial intelligence for biology. *Integrative and Comparative Biology*, 61(6), 2267-2275.
- Haudek, K. C., Prevost, L. B., Moscarella, R. A., Merrill, J., & Urban-Lurain, M. (2012). What are they thinking? Automated analysis of student writing about acid-base chemistry in introductory biology. *CBE Life Sciences Education*, 11(3), 283-293.
- Heeg, D. M., & Avraamidou, L. (2023). The use of Artificial intelligence in school science: a systematic literature review. *Educational Media International*, 60(2), 125-150.
- Hessler, G., & Baringhaus, K. H. (2018). Artificial intelligence in drug design. *Molecules*, 23(10), 2520.
- Higgins, J. P., & Green, S. (Eds.). (2008). *Cochrane handbook for systematic reviews of interventions* (Vol. 5). Wiley.
- Holmes, J., Sacchi, L., & Bellazzi, R. (2004). Artificial intelligence in medicine. *Annals of the Royal College of Surgeons of England*, 86(5), 334-338.
- Holmes, W., Bialik, M., & Fadel, C. (2023). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.
- Ijaz, K., Bogdanovych, A., & Trescak, T. (2017). Virtual worlds vs books and videos in history education. *Interactive Learning Environments*, 25(7), 904-929.
- Iyamuremye, A., Niyonzima, F. N., Mukiza, J., Twagilimana, I., Nyirahabimana, P., Nsengimana, T., ... & Nsabayeze, E. (2024). Utilization of artificial intelligence and machine learning in chemistry education: A critical review. *Discover Education*, 3(1), 95.
- Jarek, K., & Mazurek, G. (2019). Marketing and artificial intelligence. *Central European Business Review*, 8(2), 213.
- Jescovitch, L. N., Scott, E. E., Cerchiara, J. A., Merrill, J., Urban-Lurain, M., Doherty, J. H., & Haudek, K. C. (2021). Comparison of machine learning performance using analytic and holistic coding approaches across constructed response assessments aligned to a science learning progression. *Journal of Science Education and Technology*, 30(2), 150-167.
- Jho, H., & Ha, M. (2024). Towards effective argumentation: Design and implementation of a generative ai-based evaluation and feedback system. *Journal of Baltic Science Education*, 23(2), 280-291.
- Jukiewicz, M. (2024). The future of grading programming assignments in education: The role of ChatGPT in automating the assessment and feedback process. *Thinking Skills and Creativity*, 52, 101522.
- Karaçam, Z. (2013). Sistematik derleme metodolojisi: Sistematik derleme hazırlamak için bir rehber (Systematic review methodology: A Guide to preparing systematic reviews). *Dokuz Eylül Üniversitesi Hemşirelik Fakültesi Elektronik Dergisi*, 6(1), 26-33.
- Kim, N. J., & Kim, M. K. (2022). Teacher's perceptions of using an artificial intelligence-based educational tool for scientific writing. *Frontiers in Education*, 7, 755914.
- Koc-Januchta, M. M., Schönborn, K. J., Roehrig, C., Chaudhri, V. K., Tibell, L. A. E., & Heller, H. C. (2022). "Connecting concepts helps put main ideas together": Cognitive load and usability in learning biology with an AI-Enriched textbook. *International Journal of Educational Technology in Higher Education*, 19(1), 11.

- Koć-Januchta, M. M., Schönborn, K. J., Tibell, L. A. E., Chaudhri, V. K., & Heller, H. C. (2020). Engaging with biology by asking questions: Investigating students' interaction and learning with an Artificial Intelligence-Enriched textbook. *Journal of Educational Computing Research*, 58(6), 1190-1224.
- Koedinger, K. R., & Corbett, A. (2005). Cognitive tutors. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 61-78). Cambridge University Press.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.
- Lidiastuti, A. E., Herak, R., Darmawan, H., Yuniarti, N., & Yane, S. (2025). The role of artificial intelligence in enhancing biology education: A bibliometric perspective. *BIO-INOVED: Jurnal Biologi-Inovasi Pendidikan*, 7(2), 298-307.
- Lin, Y. T., & Ye, J. H. (2023). Development of an educational chatbot system for enhancing students' biology learning performance. *Journal of Internet Technology*, 24(2), 275-281.
- Lu, Q., Yao, Y., Xiao, L., Yuan, M., Wang, J., & Zhu, X. (2024). Can ChatGPT effectively complement teacher assessment of undergraduate students' academic writing?. *Assessment & Evaluation in Higher Education*, 49(5), 616-633.
- Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson.
- Martin, F., Zhuang, M., & Schaefer, D. (2024). Systematic review of research on artificial intelligence in K-12 education (2017-2022). *Computers and Education: Artificial Intelligence*, 6, 100195.
- McCarthy, J. (1987). Generality in artificial intelligence. *Communications of the ACM*, 30(12), 1030-1035.
- Miles, M. B., & Huberman, A. M. (2016). *An extended sourcebook: Qualitative data analysis*. Pegem Akademik Publishing.
- Miller, J. C., Miranda, J. P. P., & Tolentino, J. C. G. (2025). Artificial intelligence in physical education: A review. In M. B. Garcia (Ed.), *Global innovations in physical education and health* (pp. 37-60). IGI Global.
- Minsky, M., & Papert, S. A. (1969). *Perceptrons: An introduction to computational geometry*. MIT Press.
- Moharrerri, K., Ha, M., & Nehm, R. H. (2014). EvoGrader: An online formative assessment tool for automatically evaluating written evolutionary explanations. *Evolution: Education and Outreach*, 7(1), 15.
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *International Journal of Surgery*, 8(5), 336-341.
- Mubarak, A. A., Cao, H., & Zhang, W. (2022). Prediction of students' early dropout based on their interaction logs in online learning environment. *Interactive Learning Environments*, 30(8), 1414-1433.
- Nasution, N. E. A. (2023). Using artificial intelligence to create biology multiple choice questions for higher education. *Agricultural and Environmental Education*, 2(1), em002.
- Newell, A., & Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, 19(3), 113-126.
- Ouyang, F., Zheng, L., & Jiao, P. (2022). Artificial intelligence in online higher education: A systematic review of empirical research from 2011 to 2020. *Education and Information Technologies*, 27(6), 7893-7925.
- Oxford University Press. (n.d.). *Artificial intelligence*. In *Oxford English dictionary*. Retrieved from <https://www.oed.com/dictionary/artificial-intelligence>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372, 71.
- Parunak, H. V. D. (1996). Artificial intelligence in industry. In N. R. Jennings & G. M. P. O'Hare (Eds.), *Foundations of distributed artificial intelligence* (pp. 139-164). Wiley.
- Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). Cambridge University Press.
- Peffer, M. E., Ramezani, N., Quigley, D., Royse, E., & Bruce, C. (2020). Learning analytics to assess beliefs about science: Evolution of expertise as seen through biological inquiry. *CBE Life Sciences Education*, 19(3), 1-18.
- Pham, D. T., & Pham, P. T. N. (1999). Artificial intelligence in engineering. *International Journal of Machine Tools and Manufacture*, 39(6), 937-949.
- Royse, E. A., Manzanares, A. D., Wang, H., Haudek, K. C., Azzarello, C. B., Horne, L. R., Druckenbrod, D. L., Shiroda, M., Adams, S. R., Fairchild, E., Vincent, S., Anderson, S. W., & Romulo, C. (2024). FEW questions, many answers: Using machine learning to assess how students connect food-energy-water (FEW) concepts. *Humanities and Social Sciences Communications*, 11(1), 1033.
- Rusmana, A. N., Aini, R. Q., Sya'bandari, Y., Ha, M., Shin, S., & Lee, J.-K. (2021). Probing high school students' perceptions of the concept of species: A semantic network analysis approach. *Journal of Biological Education*, 55(5), 472-486.
- Sadava, D., Hillis, D. M., Heller, H. C., & Berenbaum, M. R. (2014). *Life science biology*. In E. Gündüz & I. Türkan (Eds., Trans.). Palme Publishing.
- Salas-Pilco, S. Z., Xiao, K., & Hu, X. (2022). Artificial intelligence and learning analytics in teacher education: A systematic review. *Education Sciences*, 12(8), 569.

- Sengar, S. S., Hasan, A. B., Kumar, S., & Carroll, F. (2025). Generative artificial intelligence: a systematic review and applications. *Multimedia Tools and Applications*, 84(21), 23661-23700.
- Sharadgah, T. A., & Sa'di, R. A. (2022). A systematic review of research on the use of artificial intelligence in English language teaching and learning (2015-2021): What are the current effects?. *Journal of Information Technology Education: Research*, 21, 337-377.
- Sripathi, K. N., Moscarella, R. A., Steele, M., Yoho, R., You, H., Prevost, L. B., Urban-Lurain, M., Merrill, J., & Haudek, K. C. (2024). Machine learning mixed methods text analysis: An illustration from automated scoring models of student writing in biology education. *Journal of Mixed Methods Research*, 18(1), 48-70.
- Strauss, A. L., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: Sage.
- Todd, A., Romine, W., Sadeghi, R., Cook Whitt, K., & Banerjee, T. (2022). How do high school students' genetics progression networks change due to genetics instruction and how do they stabilize years after instruction? *Journal of Research in Science Teaching*, 59(5), 779-807.
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(236), 433-460.
- Uhl, J. D., Sripathi, K. N., Meir, E., Merrill, J., Urban-Lurain, M., & Haudek, K. C. (2021). Automated writing assessments measure undergraduate learning after completion of a computer-based cellular respiration tutorial. *CBE Life Sciences Education*, 20, 33.
- Urry, L. A., Cain, M. L., Wasserman, S. A., Minorsky, P. V., & Orr, R. B. (2022). *Campbell biyoloji* (Campbell biology) In E. Gündüz & I. Türkan (Eds., Trans.). Palme Yayıncılık.
- Vater, A., Mayoral, J., Nunez-Castilla, J., Labonte, J. W., Briggs, L. A., Gray, J. J., Makarevitch, I., Rumjahn, S. M., & Siegel, J. B. (2021). Development of a broadly accessible, computationally guided biochemistry course-based undergraduate research experience. *Journal of Chemical Education*, 98(2), 400-409.
- Wang, K., Feng, Z., Li, J., & Han, R. (2019). A structural design and interaction algorithm of smart microscope embedded on virtual and real fusion technologies. *IEEE Access*, 7, 152088-152102.
- Wang, S., Wang, F., Zhu, Z., Wang, J., Tran, T., & Du, Z. (2024). Artificial intelligence in education: A systematic literature review. *Expert Systems with Applications*, 252, 124167.
- Webb, S. (2018). Deep learning for biology. *Nature*, 554(7693), 555-557.
- Xu, M., Liu, D., & Zhang, Y. (2022). Design of interactive teaching system of physical training based on artificial intelligence. *Journal of Information & Knowledge Management*, 21(Supp02), 2240021.
- Yıldırım, A. & Şimşek, H. (2016). *Qualitative research methods in social sciences*. Seckin Publishing.
- Yin, J., Goh, T.-T., & Hu, Y. (2024). Interactions with educational chatbots: The impact of induced emotions and students' learning motivation. *International Journal of Educational Technology in Higher Education*, 21, 80.
- Yin, J., Zhu, Y., Goh, T.-T., Wu, W., & Hu, Y. (2024). Using educational chatbots with metacognitive feedback to improve science learning. *Applied Sciences*, 14(20), 9345.
- Zafeiropoulos, V., & Kalles, D. (2024). Using machine learning to calibrate automated performance assessment in a virtual laboratory: Exploring the trade-off between accuracy and explainability. *Applied Sciences*, 14(17), 7944.
- Zawacki-Richter, O., Kerres, M., Bedenlier, S., Bond, M., & Buntins, K. (Eds.). (2020). *Systematic reviews in educational research: Methodology, perspectives and application*. Springer.
- Zhai, X., Chu, X., Chai, C. S., Jong, M. S. Y., Istenic, A., Spector, M., ... & Li, Y. (2021). A Review of artificial intelligence (AI) in education from 2010 to 2020. *Complexity*, 2021(1), 8812542.
- Zhang, L., & VanLehn, K. (2016). How do machine-generated questions compare to human-generated questions? *Research and Practice in Technology Enhanced Learning*, 11, 3.
- Zhang, P., & Tur, G. (2024). A systematic review of ChatGPT use in K-12 education. *European Journal of Education*, 59(2), e12599.

Author(s) Information

Tugce Duran

Department of Biology Education, Ahmet Kelesoglu
Faculty of Education, Necmettin Erbakan University,
Konya, Türkiye
Contact e-mail: tugce.gulesir@erbakan.edu.tr
ORCID iD: <https://orcid.org/0000-0003-2428-7635>

Musa Dikmenli

Department of Biology Education, Ahmet Kelesoglu
Faculty of Education, Necmettin Erbakan University,
Konya, Türkiye
ORCID iD: <https://orcid.org/0000-0001-6501-9034>
