HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

SOFTWARE AND PEDAGOGICAL FOUNDATIONS OF THE USE OF MULTIMODAL ARTIFICIAL INTELLIGENCE TECHNOLOGIES IN THE EDUCATIONAL PROCESS

Giyosova Gulnoza Alisher qizi

E-mail: giyasovagulnoza09@gmail.com

Doi: https://doi.org/10.5281/zenodo.21932884

Abstract: The article analyses the software and pedagogical foundations of using multimodal artificial intelligence (AI) technologies in the educational process. The study aims to substantiate the correspondence between the technical architecture of multimodal models and the principles of learning theory, and to identify the pedagogical conditions for their integration into teaching. The theoretical analysis describes the software basis of multimodal AI systems through a three-stage scheme: encoding of modalities, fusion in a shared semantic space, and response generation. R. Mayer's cognitive theory of multimedia learning and J. Sweller's cognitive load theory are applied as the pedagogical basis, with an interpretation of the modality, redundancy, coherence and segmenting principles under multimodal AI conditions. UNESCO's AI competency frameworks for teachers and for students are compared. A four-stage implementation model is proposed, together with five key risks and pedagogical solutions for mitigating them.

Key words: multimodal artificial intelligence, educational process, software foundation, pedagogical foundation, cognitive load, multimedia learning theory, TPACK, UNESCO competency framework, digital didactics, learning analytics.

Introduction

The entry of artificial intelligence technologies into education has reached a qualitatively new stage in recent years: systems limited to text have given way to multimodal models capable of simultaneously perceiving and analysing text, images, sound and video. This shift is not accidental for education — the learning process is itself multimodal by nature, since a learner acquires knowledge through text, image, speech and action at the same time. For this reason, from a didactic point of view, multimodal AI tools stand closer to the learning process than conventional digital tools.

In Uzbekistan this direction is supported at the level of state policy: Resolution No. PQ-358 of the President of the Republic of Uzbekistan dated 14 October 2024 approved the Strategy for the Development of Artificial Intelligence Technologies until 2030, which defines the tasks of establishing research laboratories, training qualified personnel and increasing the share of AI-based services [1]. At the international level, in 2024 UNESCO published AI competency frameworks for teachers and for students [2; 3], thereby establishing the pedagogical standards for the implementation of this technology.

At the same time, the principal problem in practice is methodological rather than technological: multimodal tools are often applied without didactic justification, following the principle of "the capability exists, so let us use it". As a result, the learner's cognitive load increases and the effectiveness of learning declines. The aim of this article is to connect the software foundations of multimodal AI technologies with the principles of pedagogical theory and to determine the scientifically grounded conditions for their integration into the educational process.

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

Research methodology

The research was carried out using the methods of theoretical analysis, comparative analysis and the systems approach. The theoretical basis was formed by R. Mayer's cognitive theory of multimedia learning [4; 5], J. Sweller's cognitive load theory [6] and the TPACK model of P. Mishra and M. Koehler [7]. UNESCO's competency frameworks [2; 3] and its guidance on generative AI [8], as well as the OECD's analytical report on digital education [9], were studied as the normative and methodological basis. The analysis was conducted along two dimensions: the software dimension (the architecture and functional capabilities of the models) and the pedagogical dimension (correspondence to didactic principles).

Results and their analysis

1. Software foundations of multimodal AI technologies. A multimodal model operates by bringing several types of data into a single internal representation. Its software basis may conditionally be divided into three stages: (a) encoding — each modality (text, image, audio, video) is converted into a vector representation through its own encoder; (b) fusion — the resulting vectors are projected into a shared semantic space and aligned with one another; (c) generation — a response in the form of text, image or sound is produced on the basis of the fused representation. It is precisely the second stage that is decisive for education, since it enables the system to assess a learner's written answer, drawing and oral explanation within a single integral context (Table 1).

Table 1

Modalities of multimodal AI and their didactic functions in the educational process

Modality	**Software capability**	**Didactic function in the educational process**
Text	natural language understanding and generation	questions and answers, text analysis, feedback, note-taking
Image	object recognition, reading diagrams and drawings	analysis of geometric drawings, maps and laboratory images
Audio / speech	converting speech into text and text into speech	pronunciation assessment, listening comprehension, inclusive support
Video	analysis of movement and sequences	observation of practical skills and the performance of experiments
Code / formula	analysis of structured languages	step-by-step checking of the solution of a problem

Source: compiled by the author on the basis of materials [8]; [9].

2. Pedagogical foundations: correspondence with cognitive theory. Multimodality in itself does not automatically improve learning. According to R. Mayer's theory, a person perceives information through two independent channels — the visual and the verbal — yet the capacity of each channel is limited [4, pp. 27–45]. J. Sweller, in turn, substantiated that an excessive flow of information increases extraneous cognitive load and impedes learning [6, pp. 37–76]. For this reason, the following principles must be observed when applying multimodal AI tools (Table 2).

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

Table 2

Interpretation of the principles of multimedia learning under multimodal AI conditions

Principle	Content	Application in multimodal AI	Risk when violated
Modality	an image with an oral explanation is more effective than one with text	presenting a diagram with audio commentary	overloading of the visual channel
Redundancy	not repeating the same information in two channels	not presenting the text both on screen and in audio	splitting of attention
Coherence	removing superfluous decoration and effects	simplifying the generated content	loss of the main idea
Segmenting	dividing the material into assimilable portions	a step-by-step response mode	excessive overloading of working memory
Signalling	highlighting the important elements	marking and explaining the key concepts	increase in independent search time

Source: compiled by the author on the basis of [4]; [5]; [6].

3. Competency foundation. The implementation of the technology depends directly on the teacher's preparedness. According to the TPACK model, effective application arises at the intersection of technological, pedagogical and content knowledge; if any one of them is lacking, the tool performs merely the function of external decoration [7, pp. 1017–1054]. In 2024 UNESCO gave this requirement a normative form: 15 competencies in five dimensions were defined for teachers and 12 competencies in four dimensions for students, each of them described at three levels (Table 3) [2; 3].

Table 3

Comparative description of the UNESCO AI competency frameworks (2024)

Parameter	Framework for teachers [2]	Framework for students [3]
Number of dimensions	5	4
Number of competencies	15	12
Levels	acquire — deepen — create	understand — apply — create

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

Parameter	Framework for teachers [2]	Framework for students [3]
Main dimensions	human-centred mindset; ethics of AI; AI foundations and applications; AI pedagogy; AI for professional development	human-centred mindset; ethics of AI; AI techniques and applications; AI system design
Practical function	a standard for teacher training programmes	a standard for the content of curricula

Source: compiled by the author on the basis of [2]; [3].

4. Implementation model and risks. As a result of the analysis, a four-stage sequence for integrating multimodal AI tools into the educational process was substantiated: (1) defining the didactic objective — determining which modality is required for which learning outcome; (2) selecting the tool and testing it; (3) configuring the feedback and assessment mechanism; (4) measuring the outcome through learning analytics. Violation of this sequence, that is, proceeding from the tool to the objective rather than from the objective to the tool, is the most frequent error in practice. The main risks arising during implementation and their pedagogical solutions are presented in Table 4.

Table 4

Main risks in the implementation of multimodal AI and pedagogical solutions

Risk	Cause	Pedagogical solution
Cognitive overload	presentation of several modalities at once	observing the principles of redundancy and segmenting
Weakening of independent thinking	use of ready-made answers	assessing the process; requiring justification of the answer
Unreliable information (hallucination)	the probabilistic nature of the generative model	including the skill of verifying sources in the learning objectives [8]
Data privacy	transfer of learner data to an external service	a procedure for working without personal data and a consent standard
Digital inequality	differences in access to devices and the internet	offline alternative assignments; use of institutional resources

Source: compiled by the author on the basis of [8]; [9]; [10].

Discussion

The analysis carried out shows that the value of multimodal AI in education is determined not by its technical capabilities but by its correspondence to the didactic objective. Increasing the number of modalities improves learning only when the structure of the learning material accords with the principles of cognitive theory; otherwise it turns into a source of additional extraneous load [4; 6]. In this sense multimodal AI does not differ from traditional multimedia learning tools — the only

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

difference is that it does not prepare content in advance but generates it in real time, which makes didactic control all the more necessary.

The second important aspect is the transformation of the assessment system. Since multimodal tools can easily produce a finished result, the centre of assessment must shift from the product to the process: what decisions the learner made at which stage, and how they verified and corrected the AI's response. Such an approach corresponds to the "human-centred" principle of the UNESCO guidance [8] and to the conclusions on the need to reconsider learning outcomes under the conditions of generative AI [10]. A limitation of the study is that its conclusions rest on theoretical analysis; experimental verification of the effectiveness of the proposed model constitutes a task for further research.

Conclusion

1. The software basis of multimodal AI systems consists of a three-stage process — encoding of modalities, fusion in a shared semantic space and response generation — and it is precisely the fusion stage that makes it possible to assess a learner's activity holistically in education.

2. The cognitive theory of multimedia learning and cognitive load theory serve as the pedagogical basis; the principles of modality, redundancy, coherence, segmenting and signalling must be taken into account in the design of multimodal AI tools.

3. It is expedient to form the teacher's preparedness on the basis of the TPACK model and UNESCO's 2024 competency frameworks (15 competencies in 5 dimensions for teachers and 12 competencies in 4 dimensions for students).

4. Implementation should be carried out in the sequence "didactic objective → tool → feedback → learning analytics", and the centre of assessment should be shifted from the product to the process; this reduces the risks of cognitive overload, the weakening of independent thinking, unreliable information, breaches of privacy and digital inequality.

References

1. Resolution No. PQ-358 of the President of the Republic of Uzbekistan dated 14 October 2024 "On approval of the Strategy for the Development of Artificial Intelligence Technologies until 2030". – Tashkent, 2024.
2. UNESCO. AI competency framework for teachers. – Paris: UNESCO, 2024. – 66 p.
3. UNESCO. AI competency framework for students. – Paris: UNESCO, 2024. – 58 p.
4. Mayer R.E. Multimedia Learning. – 3rd ed. – Cambridge: Cambridge University Press, 2021. – 372 p.
5. Mayer R.E. Cognitive theory of multimedia learning // The Cambridge Handbook of Multimedia Learning. – Cambridge: Cambridge University Press, 2014. – P. 43–71.
6. Sweller J. Cognitive load theory // Psychology of Learning and Motivation. – 2011. – Vol. 55. – P. 37–76.
7. Mishra P., Koehler M.J. Technological Pedagogical Content Knowledge: A framework for teacher knowledge // Teachers College Record. – 2006. – Vol. 108, № 6. – P. 1017–1054.

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

8. UNESCO. Guidance for generative AI in education and research. – Paris: UNESCO, 2023. – 44 p.
9. OECD. Digital Education Outlook 2023: Towards an Effective Digital Education Ecosystem. – Paris: OECD Publishing, 2023. – 194 p.
10. Baidoo-Anu D., Owusu Ansah L. Education in the era of generative artificial intelligence // Journal of AI. – 2023. – Vol. 7, № 1. – P. 52–62.