Classification of physics problems as a basis for the development of educational AI models

Milena Živković[1], Dubravka Živković[2], Marina Svičević[3], Aleksandar Milenković[3], Nemanja Vučićević[3]

[1]Department of Physcis, Faculty of Science, University of Kragujevac,Serbia, milena.zivkovic@pmf.kg.ac.rs
[2]Department of Psychology, Faculty of Philosophy, University of Niš, Serbia, dubravkazivkovic3@gmail.com
[3]Department of Mathematics and Informatics, Faculty of Science, University of Kragujevac,Serbia, marina.svicevic@pmf.kg.ac.rs, aleksandar.milenković@pmf.kg.ac.rs, nemanja.vucicevic@pmf.kg.ac.rs

The integration of artificial intelligence (AI) into education has opened new possibilities for personalized learning, adaptive assessment, and intelligent content delivery (Tapalova & Zhiyenbayeva, 2022). However, the effectiveness of such systems largely depends on the availability of high-quality, well-structured datasets that reflect the complexity of real educational content. In the domain of science education, and particularly physics, problem-solving tasks represent a core component of learning and assessment (Küchemann et al., 2024). Despite their importance, these tasks are rarely prepared in formats suitable for machine learning applications. To address this gap, we propose a structured classification framework for physics problems at the elementary school level (Shamshin, 2024; de Souza et al., 2024). Our approach focuses on the systematic annotation and organization of tasks with pedagogically and cognitively relevant features, creating a dataset suitable for training and evaluating AI models in education. By classifying physics problems based on problem type, cognitive complexity, physical quantities, and other key attributes, this framework supports the development of intelligent educational systems capable of adaptive task recommendation, personalized learning, and formative assessment. The problems were processed and annotated according to relevant criteria, including problem type (conceptual, quantitative, mixed), cognitive complexity level according to revised Bloom's taxonomy (Krathwohl, 2002), number of physical quantities and formulas, key concepts, and measurement units (Table 1). Problem complexity is determined by the number of reasoning steps required, from simple calculations (e.g., finding speed using $v = s/t$) to multi-concept problems (e.g., calculating acceleration while considering friction and inclined forces). For instance, determining a box's acceleration when pulled at an angle requires resolving forces, calculating friction, and applying Newton's laws - a 5-step analysis typical for 8th grade physics. The dataset covers key physics topics for grades 7 and 8, such as force and motion, oscillations, and optics.

Table 1. Classification Criteria for Physics Problems for AI Modeling

Category	Description
Problem Type	Quantitative, Conceptual, Mixed
Cognitive Level	Bloom's Taxonomy: Remember, Understand, Apply, Analyze
Physical Quantities	Number of physical quantities involved in the problem
Formulas	Number and complexity of formulas applied
Key Concepts	E.g., force, friction, reflection, refractive index

Units	SI units appearing in the problem
Presence of Figures	With figures, without figures
Problem Complexity	Number of steps required to reach the correct solution
Language	Serbian and English – for international use in AI applications

The classification was conducted with a focus on didactic clarity and was structured into a digital database suitable for training and evaluating AI models. This approach enables the development of systems for automatic problem classification and selection, content recommendation based on students' prior knowledge, and personalized learning. The proposed framework serves as a foundation for the development of adaptive and intelligent educational tools in physics teaching.

A prospective full-length paper should focus on the analysis and discussion of specific AI models and the ways in which these models could leverage the developed physics problem dataset. For example, transformer-based architecture could use this dataset for automated problem difficulty prediction by identifying patterns between problem features (e.g., formula count) and student performance metrics. Additionally, the dataset could be used to train reinforcement learning agents for personalized task recommendation systems that adapt the sequence of problems based on a student's mastery of individual concepts, such as force or optics. Furthermore, generative models could utilize the annotated problem structures to synthesize new, pedagogically sound physics questions while maintaining an appropriate level of complexity aligned with Bloom's taxonomy. Further extensions of this research could include comparing different Large Language Models (LLMs), such as o3 (OpenAI), Claude Opus 4 (Anthropic), DeepSeek R1, and Gemini Advanced (Google DeepMind), in the context of processing and applying this structured physics problem dataset. Such an analysis would help identify which architecture best support physics education tasks while also highlighting their limitations, such as the occurrence of hallucinations in derivations and reasoning processes. Of particular importance would be exploring the potential of fine-tuning open-source models on this specific dataset to achieve higher accuracy and reliability when solving similar types of physics problems. The results of such research could serve as valuable guidelines for the selection and adaptation of AI models in real-world educational scenarios.

Keywords: physics education, problem classification, artificial intelligence, Bloom's taxonomy, data set

References:

[1] Küchemann, S., Steinert, S., Kuhn, J., Avila, K., & Ruzika, S. (2024). Large language models—Valuable tools that require a sensitive integration into teaching and learning physics. *The physics teacher*, *62*(5), 400-402.
[2] Shamshin, A. (2024). Implementation of artificial intelligence in an online platform for solving and checking the solutions of physical problems. In *International Scientific-Practical Conference" Information Technology for Education, Science and Technics"* (pp. 400-417). Cham: Springer Nature Switzerland.
[3] de Souza, M. G., Won, M., Treagust, D., & Serrano, A. (2024). Visualising relativity: assessing high school students' understanding of complex physics concepts through AI-generated images. *Physics Education*, *59*(2), 025018.
[4] Tapalova, O., & Zhiyenbayeva, N. (2022). Artificial intelligence in education: AIEd for personalised learning pathways. *Electronic Journal of e-Learning*, *20*(5), 639-653.
[4] Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. Theory into practice, 41(4), 212-218.