

UDC 373.3:004.8(497.11)

Original scientific paper

Submitted 25/4/2025

Accepted 12/5/2025

doi:10.5937/metpra28-58483

Gordana M. Miščević¹

Gorana M. Starijaš²

Ivana M. Petrović³

University of Belgrade, Faculty of Education, Belgrade (Serbia)

ASSESSMENT OF TASKS CREATED BY CHATGPT-4o FOR 10-YEAR-OLDS IN THE FIELD OF MAGNETISM⁴

Abstract: This paper examines the potential of ChatGPT-4o as a support tool in Nature and Society teaching for fourth-grade students. Tasks of three levels of complexity in the domain of magnetism were generated by the tool and evaluated by 12 experts using a Likert scale. The results indicate that the generated tasks are age-appropriate, with the basic-level tasks receiving the highest evaluations. The analysis focused on aspects such as safety, promotion of student cooperation, alignment with learning outcomes, real-life application, and clarity. Certain shortcomings were noted, particularly in terms of safety in task implementation and the need for clearer instructional guidelines. The paper discusses the implications for future use of AI-based systems in elementary education and emphasizes the necessity of teacher training for responsible implementation of such technologies.

Keywords: ChatGPT-4o, artificial intelligence in education, natural sciences, magnetism, teacher training, elementary school students

1 gordana.miscevic@uf.bg.ac.rs, <https://orcid.org/0000-0002-7668-7596>; year of birth 1974.

2 gorana.starijas@uf.bg.ac.rs, <https://orcid.org/0009-0003-7370-6661>; year of birth 1982.

3 ivana.petrovic@uf.bg.ac.rs, <https://orcid.org/0000-0002-5571-0945>; year of birth 1992.

4 This research was conducted by the University of Belgrade – Faculty of Education, as part of the project „Building the Critical Computer Skills for the Future-Ready Workforce” (No. 00136459) implemented by the United Nations Development Programme (UNDP) in partnership with the Ministry of Education, with the support of the Government of the Republic of Serbia.

INTRODUCTION

Instruction based on traditional teaching methods, the excessive use of frontal teaching, and the passive role of students in the learning process can no longer meet the needs of today's learners (Hu, Wang, Wang, 2024; Mandić, 2023). For this reason, educators—particularly those involved in teacher education—advocate for innovative instructional approaches that lead more effectively and efficiently to the development of students' competencies. One of the most current innovations involves the integration of artificial intelligence (AI) into the educational process. While the first chatbot prototypes appeared in the previous century—Eliza in 1966, Parry in 1972, and Alice in the mid-1990s—the release of OpenAI's ChatGPT in November 2022 marked a turning point. In a very short time, this tool reached a vast number of users and entered the field of education. Various studies have identified several factors that influence the use of chatbots in education, including perceived usefulness, ease of use, positive attitudes toward AI chatbots, and expected academic benefits. Interestingly, academic workload, time pressure, and reward sensitivity are also noted as significant factors (Abbas, Jam, Khan, 2024). Chatbots have shown potential for simplifying complex concepts, thus enhancing understanding (Shuhaiber et al., 2025). In light of these recent changes, the necessity for acquiring fundamental AI-related knowledge and skills has been emphasized even from the preschool level (Kölemen, Yıldırım, 2025). For example, parents in Hong Kong view the use of AI in kindergartens positively (Su, 2025). It is possible to foster algorithmic thinking through play-based learning and the use of AI robots. Research has involved the use of tutor robots, capable of physical movement and emotional recognition. These robots interact with students to personalize educational content by utilizing both psychological and cognitive indicators (Maaz, Mounsef, Maalouf, 2025). Researchers are developing various methodological approaches to implement educational robots as tools for developing algorithmic thinking in preschool children (Matović & Ristić, 2024). For these ideas to be successfully implemented, it is essential to investigate preschool teachers' conceptual understanding of the topic. Although preservice preschool teachers generally have a positive attitude toward the potential of educational robots, research shows that misconceptions remain due to underdeveloped digital competencies (Mandić, Miščević, Babić, Matović, 2024). Preservice preschool teachers and preservice elementary school teachers have experience using chatbots, but those experiences are not typically acquired during their academic studies. Therefore, research of this kind provides valuable guidance for improving formal education programs at the institutional level (Mandić, Miščević, Bujišić, 2024).

PRESENCE OF THE TOPIC OF MAGNETISM IN PRIMARY EDUCATION

Experiments in which the phenomenon of magnetism is revealed through the game method are often represented even in the preschool period, because the development of scientific competences is one of the key competencies of the preschool program in Serbia (Fundamentals of the Preschool Education Program Godine uzleta, 2018). In the Republic of Serbia, magnetism is formal introduced for the first

time in the fourth grade of primary school as part of the school subject *Nature and Society*, within the broader topic of *Materials* (Program for the Fourth Grade of Primary Education, 2019). The program content focuses on investigating magnetic properties of materials (natural magnets, the possibility of magnetizing objects, and the resulting specificities), while the expected learning outcomes state that by the end of the grade, students should be able to identify examples of magnet use in everyday life.

This topic can be enriched through excursions that help students connect theoretical content with real-life experiences (Vučetić & Andđelković, 2024). An analysis of the curricula from the first three grades of elementary school reveals a gradual introduction to the basics of materials and their properties, laying the foundation for understanding magnetism in the fourth grade. These prior learning experiences allow students to expand their knowledge and apply newly learned concepts related to magnetic specificities in everyday situations.

The topic is further developed in more depth in the sixth grade within the *Physics* curriculum (Program for the Sixth Grade of Primary Education, 2019). At this level, students explore magnetic attraction and repulsion, magnetic forces, and their connection to electrical phenomena. The foundational concepts introduced in the fourth grade are expanded with content on electrical charge, conductivity, and magnetic properties of materials. In the eighth grade, an entire unit is devoted to *Magnetic Fields* (Program for the Eighth Grade of Primary Education, 2019), in which students learn about magnetic fields of permanent magnets and the Earth, magnetic fields generated by electric current, and the contributions of Nikola Tesla and Mihajlo Pupin to the science and technological applications of electromagnetism. At the end of primary education, students perform demonstrative experiments in the domain of magnetism.

From this overview, we can conclude that science concepts in the primary curriculum are cyclically built from grade to grade. Existing knowledge is continuously connected with new information, enabling students to construct a cohesive knowledge framework on a given topic. This is also evident in the case of magnetism. In the first educational cycle, students become familiar with the basic properties of materials; later, in the second cycle, they explore magnetism and eventually integrate this understanding with abstract concepts such as magnetic fields and their relation to electric currents and technological applications. To build a strong foundation for later physics education, instructional activities for ten-year-old students should include a range of task types—from simple inquiry-based activities to more complex experiments.

METHODOLOGY

As previously mentioned, magnetism is a topic included in the fourth-grade *Nature and Society* program and later is developed in greater depth within the *Physics* curriculum. The focus of this research was to collect and analyze the assessments of experts—university professors and teaching assistants specializing in the methodology of *Nature and Society*—regarding the alignment of AI-generated tasks with educational standards for ten-year-old students, as well as to explore the potential of artificial intelligence in enabling innovative and engaging learning experiences.

The aim of the study was to provide an expert evaluation of the adequacy of the responses generated by artificial intelligence (ChatGPT-4o version) to two predefined prompts.

The evaluation instrument (Appendix 1) required experts to assess ChatGPT-4o's responses to the following two prompts:

1. "Create engaging and age-appropriate tasks on magnetism for 10-year-old students."
2. "Create three groups of magnetism tasks for 10-year-old students, categorized by difficulty level."

For both prompts, participants—experts in the methodology of teaching Nature and Society—were asked to assess the tasks using a five-point Likert scale, along with a section for descriptive comments. For the first prompt, the experts evaluated the age appropriateness of the generated tasks, the safety of the proposed activities, the potential for fostering student cooperation, and the alignment of the tasks with the learning outcomes defined by the official curriculum for the fourth-grade Nature and Society subject.

For the second prompt, the experts evaluated each of the three groups of tasks separately—basic, intermediate, and advanced—by judging whether the tasks were appropriately classified by level of difficulty, whether they were safe, whether they supported student collaboration, and whether they were aligned with the prescribed educational outcomes for the subject.

RESULTS

The sample consisted of 12 respondents-experts, including five university professors specializing in the methodology of Nature and Society teaching, employed at higher education institutions in Serbia that focus on teacher education, as well as seven of their teaching assistants. There are no more than 20 such narrowly specialized experts in that area for 10-year-old students at all teacher education faculties in Serbia.

The average ratings provided by the respondents for the first prompt — "Create engaging and age-appropriate tasks on magnetism for 10-year-old students" — are presented in **Table 1** (Appendix 2).

Table 1: Expert evaluation of the adequacy of tasks generated by artificial intelligence for ten-year-old students on the topic of magnetism (Prompt 1)

The average values for the second prompt — "Create three groups of magnetism tasks for 10-year-old students, categorized by difficulty level" — are shown in **Table 2** (Appendix 3).

Table 2: Expert evaluation of the adequacy of tasks in the second prompt categorized by three levels of complexity – basic, intermediate, and advanced.

Descriptive responses related to this prompt, which provide deeper insight into the reasoning behind the scores, are presented in **Appendix 4**.

DISCUSSION

In general, all expert evaluations of the ChatGPT-4o version, studied in our research, were above 3.5 on a five-point Likert scale. The lowest average ratings were given for the categories of **safety** (3.8) and **real-life applicability** (3.7), while the highest-rated category was the **age-appropriateness** of the formulated tasks (4.5).

In a previous study analyzing an earlier version of the tool (ChatGPT 3.5), it was found that the chatbot could produce inaccurate or potentially unsafe recommendations, which required expert verification of the content (Oviedo-Trespalacios et al., 2023). It is evident that newer versions produce more accurate and higher-quality responses; however, methodological design by experts remains necessary—especially with regard to the safe implementation of proposed tasks. Our findings confirm that participants agree the proposed tasks are well-aligned with students' developmental level (4.5), and that the classification of tasks into different levels of complexity is largely appropriate. Nonetheless, the participants identified a certain lack of precision in the differentiation of task complexity levels. The overall quality of the tasks was rated highest for those at the basic level of complexity. This trend is consistent across all evaluated dimensions: alignment with complexity level, clarity, cooperation potential, outcome alignment, real-life relevance, and safety—where basic-level tasks received the highest average score (4.9), compared to the other two levels (4.3). Recent research on the use of AI—including ChatGPT—in kindergartens has demonstrated the potential for generating illustrated narratives on magnets and increasing children's engagement and creativity (Samara & Kotsis, 2024).

In terms of task safety, experts emphasized the importance of supervision and detailed guidelines for conducting experiments, especially when tasks involve sharp objects or materials that could pose health risks if inhaled or mishandled. When comparing average scores across complexity levels, we observe that safety ratings tend to decrease slightly as task complexity increases. This trend likely reflects the increased need for teacher involvement and oversight in facilitating more complex, hands-on learning activities. This finding underscores another important aspect of the teacher's role: ensuring the safety of AI-generated tasks, particularly for advanced-level activities. It is essential to enhance safety by requiring teachers to analyze and, if necessary, modify AI-generated tasks to include specific guidance for safe handling of sharp or potentially hazardous materials.

The Faculty of Education at the University of Belgrade (Serbia) houses the *Center for Robotics and Artificial Intelligence in Education (CRAIE)*⁵, where virtual experiments can be conducted in a completely safe environment using the Virtual Lab platform⁶—integrating advanced 3D technologies, virtual reality, and artificial intelligence.

The possibility of fostering student cooperation through ChatGPT-generated tasks was evaluated with an average score of 4.0 for Prompt 1, and 3.9 across all three complexity levels. These results suggest that respondents recognize ChatGPT as a potentially useful tool for encouraging peer collaboration, but only when used in conjunction with appropriate teacher guidance tailored to the specific needs and

5 <https://craie.edu.rs/>

6 <https://www.youtube.com/watch?v=sx8g0Z0uE1o>

composition of the student group. Items referring to the applicability of knowledge about magnetism in everyday life received somewhat lower average scores, ranging from 3.5 to 4.1. The lowest overall average score in the instrument was recorded in this item (except for the advanced-level tasks), suggesting a need to improve activities that link scientific content to real-life situations.

CONCLUSION

The findings of our research confirm that artificial intelligence, specifically in the form of chatbots such as ChatGPT-4o, can represent a valuable pedagogical resource in planning and implementing teaching activities. Through expert evaluation of tasks in the domain of magnetism generated by this tool—conducted by university professors and assistants specializing in the methodology of teaching Nature and Society—it was confirmed that the tasks are largely aligned with the developmental characteristics of ten-year-old students, as well as with the curricular content of the subject “Nature and Society.” Although tasks at the basic level of complexity received the highest ratings across nearly all evaluated categories—including clarity, potential for student cooperation, safety, and real-life applicability—the study also highlighted several challenges in the implementation of AI tools in education. The most prominent concern relates to the safety of proposed activities and the need for explicit pedagogical instructions, particularly in the case of more complex experiments. This underscores the key role of the teacher in evaluating and methodologically shaping AI-generated tasks. To optimize the usefulness of ChatGPT in task creation, it is necessary for teachers to provide more precise guidelines concerning students’ abilities and the intended complexity of the tasks. Additionally, integrating AI-generated tasks into a blended learning environment—where AI complements rather than replaces traditional teaching methods—can help bridge gaps in both linguistic and conceptual task design. Teachers should approach AI-generated content as a supplement to, not a replacement for, their own expertise in creating personalized and age-appropriate learning activities. Special attention should also be paid to ensuring the safety of proposed activities. The expert evaluation revealed that ChatGPT-generated tasks are not fully oriented toward facilitating knowledge transfer to real-life contexts, nor do they sufficiently support the development of higher-order cognitive skills such as critical thinking and creativity. These limitations can be addressed by embedding the tasks within a broader pedagogical framework, enriched through real-life applications, interdisciplinary approaches, and problem-based learning.

Finally, the results point to the necessity of systemic training for preservice teachers in the educational use of artificial intelligence. Emphasis should be placed on developing both digital and methodological competencies, which are essential for the responsible and meaningful use of AI tools in the classroom. Institutions such as the Center for Robotics and Artificial Intelligence in Education (CRAIE) provide a strong institutional foundation for such training and represent significant potential for improving pedagogical practice in the digital age. In this context, artificial intelligence should be viewed as a supplementary tool in the hands of a reflective and professionally competent teacher—one capable of directing technology toward the development of functional knowledge, ensuring student safety, and achieving educational goals aligned with students’ age and individual learning needs.

REFERENCES

Abbas, M., Jam, F. A. & Khan, T. I. (2024). Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students. *International Journal of Educational Technology in Higher Education*, 21 (1), 10. <https://doi.org/10.1186/s41239-024-00444-7>

Fundamentals of the Preschool Education Program – Godine uzleta (2018). *Official Gazette of the Republic of Serbia*, 16, retrieved 22. 4. 2025. from

Hu, S. G., Wang, W. Y., Wang, X. X. & Yin, Y. M. (2024). Assessing the intention to accept inquiry-based teaching pedagogy among Chinese university students: an extension of technology acceptance model. *Frontiers in Psychology*, 15, 1265047. <https://doi.org/10.3389/fpsyg.2024.1265047>

Kölemen, E. B. & Yıldırım, B. (2025). A new era in early childhood education (ECE): Teachers' opinions on the application of artificial intelligence. *Education and Information Technologies*, 1–42. <https://doi.org/10.1007/s10639-025-13478-9>

Maaz, N., Mounsef, J. & Maalouf, N. (2025). CARE: towards customized assistive robot-based education. *Frontiers in Robotics and AI*, 12, 1474741. <https://doi.org/10.3389/frobt.2025.1474741>

Mandic, D. (2023). Report on Smart Education in the Republic of Serbia. In: *Smart Education in China and Central & Eastern European Countries*, 271–291. Singapore: Springer Nature Singapore

Mandić, D., Miščević, G., Babić, J. & Matović, S. (2024). Educational robots in teachers' education. *Istraživanja u pedagogiji*, 14 (2), 361–376. <https://doi.org/10.5937/IstrPed2402361M>

Mandić, D., Miščević, G. & Bujišić, Lj. (2024). Evaluating the quality of responses generated by ChatGPT. *Metodička teorija i praksa*, 27 (1), 5–19. <https://doi.org/10.5937/metpra27-51446>

Matović, S. (2024). Perception of (future) Preschool teachers on educational robots in working with children. *Istraživanja u pedagogiji*, 14 (2), 278–291. <https://doi.org/10.5937/IstrPed2402278M>

Matović, S. M. & Ristić, M. R. (2024). Educational robots as a tool for developing algorithmic thinking in preschool children. *Metodička teorija i praksa*, 27 (2), 21–35. <https://doi.org/10.5937/metpra27-52258>

Oviedo-Trespalacios, O. et al. (2023). The risks of using ChatGPT to obtain common safety-related information and advice. *Safety Science*, 167, 106244. <https://doi.org/10.1016/j.ssci.2023.106244>

Program for the Fourth Grade of Primary Education (2019). (p. 42). Retrieved 25. 3. 2025. from <https://pravno-informacioni-sistem.rs/slglrsViewPdf/7f60db16-8760-4e25-bc8a-f69fe1fb2e2c?fromLink=true>

Program for the Fourth Sixth Grade of Primary Education (2019). (p. 316). Retrieved 25. 3. 2025. from

Program for the Eighth Grade of Primary Education (2019). (p.125). Retrieved 25. 3. 2025. from <https://pravno-informacioni-sistem.rs/slglrsViewPdf/7f60db16-8760-4e25-bc8a-f69fe1fb2e2c?fromLink=true>

Rulebook on the teaching and learning program for the fourth grade of primary education (2019). (p. 42). Retrieved 25. 3. 2025. from <https://pravno-informacioni-sistem.rs/slglrsViewPdf/7f60db16-8760-4e25-bc8a-f69fe1fb2e2c?fromLink=true>

Samara, V. & Kotsis, K. T. (2024). Use of the artificial intelligence in teaching the concept of magnetism in preschool education. *Journal of Digital Educational Technology*, 4 (2), ep2419. <https://doi.org/10.30935/jdet/14864>

Shuhaiber, A., Kuhail, M. A. & Salman, S. (2025). ChatGPT in higher education - A Student's perspective. *Computers in Human Behavior Reports*, 17. <https://doi.org/10.1016/j.chbr.2024.100565>

Su, J. (2025). Kindergarten parents' perceptions of the use of AI technologies and AI literacy education: Positive views but practical concerns. *Education and Information Technologies*, 30, 279–295. <https://doi.org/10.1007/s10639-024-12673-4>

Vuletić, J. B. & Andđelković, S. D. (2024). Students' perception of the educational role of field trips in Geography education. *Metodička teorija i praksa*, 27 (2), 7–20. <https://doi.org/10.5937/met-pra27-54549>

Gordana M. Miščević⁷

Gorana M. Starijaš⁸

Ivana M. Petrović⁹

Univerzitet u Beogradu, Fakultet za obrazovanje učitelja
i vaspitača, Beograd (Srbija)

PROCENA KVALITETA ZADATAKA KOJE JE KREIRAO CHATGPT-4O ZA DESETOGODIŠNJAKE U OBLASTI MAGNETIZMA¹⁰

Sažetak: ChatGPT (Chat Generative Pre-trained Transformer) je alatka koja nudi brzo dobijanje željenih informacija, pomoći pri rešavanju problema i uz pravilnu i vođenu primenu doprinosi razvoju kritičkog mišljenja. U ovom radu istražićemo potencijal ChatGPT-4o kao inovativnog, interaktivnog alata za podršku učenju za učenike četvrtog razreda osnovne škole u oblasti nastave prirode i društva. Na osnovu postavljenog upita generisani su zadaci na tri nivoa složenosti iz oblasti magnetizma u okviru nastavnog predmeta Priroda i društvo. Cilj sprovedenog

⁷ gordana.miscevic@uf.bg.ac.rs, <https://orcid.org/0000-0002-7668-7596>; rođena 1974.

⁸ gorana.starijas@uf.bg.ac.rs, <https://orcid.org/0009-0003-7370-6661>; rođena 1982.

⁹ ivana.petrovic@uf.bg.ac.rs, <https://orcid.org/0000-0002-5571-0945>; rođena 1992.

¹⁰ Ovo istraživanje je sproveo Univerzitet u Beogradu – Fakultet za obrazovanje učitelja i vaspitača, u okviru projekta „Izgradnja kritičnih kompjuterskih veština za radnu snagu spremnu za budućnost“ (br. 00136459), koji sprovodi Program Ujedinjenih nacija za razvoj (UNDP) u partnerstvu sa Ministarstvom prosvete, uz podršku Vlade Republike Srbije.

istraživanja bila je ekspertska procena adekvatnosti odgovora koje je pružila veštačka inteligencija (ChatGPT-4o verzija) u odnosu na dva postavljena pitanja, sa smernicama za obuku budućih učitelja za njeno adekvatno korišćenje u svom budućem radu sa decom mlađeg osnovnoškolskog uzrasta. Dobijeni zadaci su vrednovani putem instrumenta koji je sačinjen od petostepene skale Likertovog tipa od strane stručnjaka iz oblasti Metodike nastave prirode i društva. Uzorak ispitanika činilo je 12 eksperata – pet univerzitetskih profesora i sedam asistenata, metodičara nastave prirode i društva koji su zaposleni na visokoškolskim ustanovama u Srbiji koje se bave obrazovanjem učitelja. Rezultati našeg istraživanja ukazuju na to da su ispitanici saglasni da su predloženi zadaci prilagođeni uzrastu učenika. Kvalitet formulisanih zadataka je najbolje ocenjen za zadatke na osnovnom nivou složenosti i taj trend se vidi u svim analiziranim aspektima – od formulacije zadatka u skladu sa nivoom složenosti, jasnoćom, preko mogućnosti za razvoj kooperacije, ostvarivanja propisanih ishoda, primene u svakodnevnom životu, do aspekta bezbednosti. Ispitanici su tvrdnju koja se odnosi na bezbednost predloženih aktivnosti procenili nešto nižim vrednostima, posebno naglašavajući nastavnikovo vođenje aktivnosti od kog u velikoj meri zavisi njihova bezbednost, kao i pažnja usmerena ka adekvatnom čuvanju i tretiranju materijala kojima se realizuju eksperimenti. Možemo da zaključimo da učitelji u velikoj meri mogu da koriste ovaj alat kao pomoćno sredstvo prilikom formulisanja zahteva za aktivnosti učenika uz precizne smernice za prilogođavanje različitim nivoima složenosti uz potrebne informacije kao što su: usmerenost na ishode učenja, prethodna znanja učenika, integracija sadržaja i sl. Neophodno je da učitelji budu svesni izazova koje sa sobom nosi primena veštačke inteligencije kako bi u najvećoj meri iskoristili prednosti ovog alata uz adekvatno upravljanje rizicima i nedostacima.

Ključне reči: ChatGPT-4o, veštačka inteligencija u obrazovanju, prirodne nauke, magnetizam, obrazovanje učitelja, učenici osnovne škole

APPENDIX

Appendix 1

Instrument

We used ChatGPT4o to create tasks for 10-year-old students about magnetism. We present to you Query 1 with received answers, You need to make an assessment of those answers. After that you have Query 2 with received answers within three groups. You need to make an assessment of those answers, too.

Query 1:

“Create engaging and age-appropriate tasks on magnetism for 10-year-old students.”

ChatGPT4o answers:

1. Magnetic or Not?

- Task: Collect a variety of objects (e.g., paperclip, rubber band, coin, plastic spoon, nail). Test each object with a magnet to see if it is magnetic or not. Create a chart to record your findings.
- Goal: Understand which materials are attracted to magnets and which are not.

2. Magnetic Maze

- Task: Use a magnet to guide a paperclip through a maze drawn on a piece of cardboard or paper. The magnet must stay underneath the surface while you move the paperclip above it.
- Goal: Explore how magnetic forces work through materials.

3. Magnetic Strength Test

- Task: Use a magnet to pick up as many paperclips as possible. Try with different magnets and compare their strength by counting the number of paperclips each magnet can hold.
- Goal: Investigate the strength of different magnets.

4. Make a Magnetic Compass

- Task: Magnetize a needle by rubbing it with a magnet in one direction multiple times. Float the needle on a small piece of cork in a bowl of water and observe how it points north.
- Goal: Learn how magnetism is used in navigation.

5. Magnetic Field Art

- Task: Place a magnet under a piece of paper. Sprinkle iron filings over the paper and gently tap it. Observe the patterns formed by the magnetic field.
- Goal: Visualize magnetic fields.

6. Magnetic Attraction Race

- Task: Place a small magnetic object (like a paperclip) on a table. Use a magnet to “race”

Query 1:

“Create engaging and age-appropriate tasks on magnetism for 10-year-old students.”

The proposed activities should be evaluated in the following categories on a scale of 1 to 5

	1	2	3	4	5	Describe your assessment
Adequacy of the tasks to the age of the student	Completely inadequate	Partially inadequate	Medium fits	Partially adequate	Completely adequate	
Safety of the proposed activities	Completely unsafe	Partially unsafe	Medium level of safety	Mostly safe	Completely safe	
The existence of the possibility to develop student cooperation	Extremely poor	Poor	Average	Good	Fantastic	
Orientation of the tasks towards achieving the outcomes prescribed by the official program in the fourth grade of primary school for school subject Nature and Society	Doesn't exist at all	Exists to a very small extent	Medium exist	Partially exist	Completely exist	

Query 2:

“Create three groups of magnetism tasks for 10-year-old students, categorized by difficulty level.”

ChatGPT4o answers:**Group 1: Basic Tasks (Introduction to Magnetism)**

These tasks are straightforward and focus on exploring basic magnetic properties.

1. Magnetic or Not?

- Test objects (e.g., paperclip, plastic spoon) with a magnet to identify which are magnetic.
- Record findings in a chart.

2. Magnetic Sorting Machine

- Mix magnetic and non-magnetic objects (e.g., coins, buttons) and use a magnet to separate them.

3. Magnet Matching Game

- Match magnets by their poles (north to south, north to north). Observe what happens when similar poles face each other.

Group 2: Intermediate Tasks (Hands-On Exploration)

These tasks require observation and introduce concepts like magnetic fields and force.

1. Magnetic Strength Test

- Compare how many paperclips different magnets can lift.
- Discuss which magnet is the strongest and why.

2. Magnetic Maze

- Move a paperclip through a maze drawn on cardboard using a magnet beneath the surface.

3. Magnetic Field Art

- Sprinkle iron filings on a piece of paper with a magnet underneath to visualize the magnetic field.

Group 3: Advanced Tasks (Application and Critical Thinking)

These tasks involve creativity and applying knowledge to solve problems.

1. Make a Magnetic Compass

- Magnetize a needle and float it on a cork in water. Observe how it points north.

2. Building a Magnet-Powered Car

- Attach a magnet to a toy car and use another magnet to push or pull it without touching.

3. Magnetic Repulsion Experiment

- Hold two magnets close together with the same poles facing each other. Explore how repulsion works and explain the observation.

Query 2:

“Create three groups of magnetism tasks for 10-year-old students, categorized by difficulty level.”
The proposed activities should be evaluated in the following categories on a scale of 1 to 5

Group 1	The tasks within Group1 are Basic Tasks	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Assesment
	Safety of the proposed activities within Group1	Completely unsafe	Partially unsafe	Medium level of safety	Mostly safe	Completely safe	
	The existence of the possibility to develop student cooperation within Group1	Extremely poor	Poor	Average	Good	Fantastic	
	Orientation of the tasks within Group1 towards achieving the outcomes prescribed by the official program in the fourth grade of primary school for school subject Nature and Society	Doesn't exist at all	Exists to a very small extent	Medium exist	Partially exist	Completely exist	

Group 2	The tasks within Group2 are Intermediate Tasks	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	
	Safety of the proposed activities within Group2	Completely unsafe	Partially unsafe	Medium level of safety	Mostly safe	Completely safe	
	The existence of the possibility to develop student cooperation within Group2	Extremely poor	Poor	Average	Good	Fantastic	
	Orientation of the tasks within Group2 towards achieving the outcomes prescribed by the official program in the fourth grade of primary school for school subject Nature and Society	Doesn't exist at all	Exists to a very small extent	Medium exist	Partially exist	Completely exist	
	The tasks within Group3 are Advanced Tasks	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	
Group 3	Safety of the proposed activities within Group3	Completely unsafe	Partially unsafe	Medium level of safety	Mostly safe	Completely safe	
	The existence of the possibility to develop student cooperation within Group3	Extremely poor	Poor	Average	Good	Fantastic	
	Orientation of the tasks within Group3 towards achieving the outcomes prescribed by the official program in the fourth grade of primary school for school subject Nature and Society	Doesn't exist at all	Exists to a very small extent	Medium exist	Partially exist	Completely exist	
	Describe your comments if you have them for both queries. If you have any additional ideas or suggestions, please write them here.						

Appendix 2

Table 1: Expert Evaluation of the Adequacy of Tasks Generated by Artificial Intelligence for Ten-Year-Old Pupils – Prompt 1 (Magnetism Topic)

Statement	Mean Score (M)
Appropriateness of tasks for students' age	4.5
Safety of proposed activities	3.8
Potential for developing cooperation among students	4.0
Alignment of tasks with outcomes prescribed by the Official program for Nature and Society in the 4th grade	4.1
Emphasis on the application of magnetism in everyday life to develop functional knowledge	3.7
Clearly formulated tasks guide students toward specific conclusions in the domain of magnetism	4.2

Appendix 3

Table 2: Expert Evaluation of Tasks from Prompt 2 Categorized by Three Levels of Complexity – Basic, Intermediate, and Advanced

Statement	Basic Level	Intermediate Level	Advanced Level
Tasks correspond to the stated complexity level	4.5	3.9	4.0
Safety of the proposed activities	4.9	4.3	4.3
Potential for developing cooperation among students	3.9	3.9	3.9
Alignment with prescribed outcomes from the Nature and Society curriculum	4.5	3.6	4.1
Emphasis on real-life application of magnetism to develop functional knowledge	3.9	3.5	4.1
Clearly formulated tasks guide students to draw specific conclusions related to magnetism	4.8	4.1	4.3

Appendix 4

Descriptive Comments by experts explaining their ratings

Age Appropriateness:

The tasks are appropriate for the students' age; however, it is necessary to clarify the items used in testing (e.g., some coins are not attracted by magnets), in order to derive accurate conclusions.

Safety of Proposed Activities:

There is a need to ensure a higher level of safety in planned activities. For example, there is a risk that a student may inhale spilled iron filings or accidentally injure themselves or others. Safety could be improved by using sealed transparent plastic containers for iron filings. The proposed activities are generally safe with adequate supervision. It would be beneficial to include guidelines and notes suggesting that teachers/parents help in obtaining required materials, particularly those with sharp edges. Additional emphasis should be placed on caution when handling sharp objects such as nails, clips, and needles.

Potential for Student Cooperation:

While cooperation is assumed, it would be beneficial for instructions to explicitly suggest conducting the experiments in pairs or groups. There is a lack of specific guidance on individual roles and responsibilities. The tasks should be supported by instructions promoting mutual discussion, exchange of observations and ideas, and reasoned critique.

Alignment with Prescribed Outcomes:

Some concepts included in the tasks, such as magnetic fields or the “strength” of magnets and the spread of magnetic forces through different materials, are not part of the fourth-grade curriculum. The tasks mostly align with learning outcomes for this area but should include tasks asking students to identify specific properties of magnets after observing magnetic phenomena. There are no tasks that encourage students to conduct inquiry in their own environments—for example, testing with a refrigerator magnet at home or using a compass at school. Additionally, the tasks do not guide students to understand when magnetic poles attract and when they repel each other. The concept of a magnetic field is not officially introduced at this grade level. A major criticism is that many examples are nearly identical to those found in standard Nature and Society textbooks, indicating repetition of already familiar experiments and exercises (reproduction).

Application in Everyday Life:

There is no operationalization of the knowledge gained from experiments, nor transfer to real-life contexts (e.g., areas of human activity where this knowledge applies—construction, navigation, medicine, etc). There is also no connection between the experiment and Earth's magnetic properties, which would be essential in the first task. **Critical thinking, creativity, and knowledge application**

were mostly absent. These aspects are generally lacking due to the absence of additional supporting questions or tasks. At this age, students are expected to relate concepts to real-life examples and assess the impact of human activity on nature, both positive and negative, and to consider potential future consequences. They should be able to design simple experiments to reach conclusions and explain cause-and-effect relationships in specific natural processes.

Clarity of Task Formulation:

The tasks are mostly clearly formulated. However, in the final activity (magnetic race), it is unclear what happens if the paperclip sticks to the magnet—should the student return to the start or continue? Additionally, distance affects magnetic force. The effect of shape and size of different magnets is not specified.