Artificial Intelligence and Cybersecurity

Engineer of XXI Century'2025

A unified decision-making framework for LLM selection: bridging industrial and educational applications

Anatolii Balyk [1,*] Vasyl Oleksiuk [1,2]

[1] *Ternopil Volodymyr Hnatiuk National Pedagogical University, 2 M. Kryvonosa Str., Ternopil, 46027, Ukraine, vodinn@tnpu.edu.ua*
[2] *Institute for Digitalisation of Education of the NAES of Ukraine, 9 M. Berlynskoho Str., Kyiv, 04060, Ukraine, oleksyuk@fizmat.tnpu.edu.ua*
* *Anatolii Balyk, vodinn@tnpu.edu.ua*

Abstract: The integration of Large Language Models (LLMs) into professional fields requires a new generation of specialists capable of making informed engineering decisions about their selection and application. This paper proposes a unified decision-making framework designed to bridge the gap between industrial system evaluation and professional competency development in informatics teacher education. The framework's core is a quantitative multi-criteria model for assessing LLMs based on performance, cost, privacy, and customizability. We demonstrate its dual applicability through two distinct use cases. The first, in industrial predictive maintenance, shows how the framework guides the selection of an optimal open-source LLM for a mission-critical task. The second, in teacher training, illustrates how the same framework is used as a pedagogical tool to develop systems thinking and AI literacy. The discussion highlights the synergy created by this unified approach, fostering a direct knowledge transfer pipeline between industry needs and educational outcomes.

Keywords: large language models; multi-criteria decision-making; engineering education; teacher training; knowledge transfer; unified framework; AI literacy; ethical AI.

Zintegrowane ramy inżynieryjno-dydaktyczne do wyboru dużych modeli językowych: przypadek kształtowania kompetencji przyszłych nauczycieli informatyki

Anatolii Balyk [1,*] Vasyl Oleksiuk [1,2]

[1] *Tarnopolski Narodowy Uniwersytet Pedagogiczny im. W. Hnatiuka, 46027, Tarnopol, Ukraina, vodinn@tnpu.edu.ua*
[2] *Instytut Cyfryzacji Edukacji Narodowej Akademii Nauk Pedagogicznych Ukrainy, ul. M. Berłyńskiego 9, Kijów, 04060, Ukraina, oleksyuk@fizmat.tnpu.edu.ua*
* *Anatolii Balyk, vodinn@tnpu.edu.ua*

Streszczenie: Integracja dużych modeli językowych (LLM) w różnych dziedzinach zawodowych wymaga nowego pokolenia specjalistów, którzy potrafią podejmować świadome decyzje inżynierskie dotyczące ich doboru i zastosowania. Niniejszy artykuł przedstawia zintegrowane ramy decyzyjne, mające na celu połączenie oceny systemów przemysłowych z rozwijaniem kompetencji zawodowych w kształceniu nauczycieli informatyki. Rdzeniem ram jest ilościowy model wielokryterialny, oceniający modele LLM pod względem wydajności, kosztów, prywatności i możliwości dostosowania. Jego podwójne zastosowanie ukazano na dwóch przykładach. Pierwszy dotyczy konserwacji predykcyjnej w przemyśle i pokazuje, jak ramy wspomagają wybór najlepszego modelu open-source do realizacji krytycznego zadania. Drugi przykład, w edukacji nauczycieli, pokazuje, jak te same ramy mogą być wykorzystane jako narzędzie dydaktyczne do rozwijania myślenia systemowego i kompetencji w zakresie sztucznej inteligencji. Dyskusja podkreśla synergię wynikającą z tego zintegrowanego podejścia, które tworzy bezpośredni kanał transferu wiedzy między potrzebami przemysłu a wynikami kształcenia.

DOI: https://doi.org/10.53052/9788367652520.18

Słowa kluczowe: duże modele językowe; wielokryterialne podejmowanie decyzji; edukacja inżynierska; szkolenie nauczycieli; transfer wiedzy; zintegrowane ramy; alfabetyzacja AI; etyka sztucznej inteligencji.

1. Introduction

The era of Large Language Models (LLMs) continues to accelerate, opening new horizons for industrial and educational engineering. The capabilities demonstrated by foundational models such as GPT-4 [1], whose emergent abilities were described as "sparks of artificial general intelligence," have evolved rapidly. This has been followed by the release of even more powerful systems, including OpenAI's GPT-4o [2], Anthropic's Claude 3 [3], Meta's Llama 3 [4], and Google's Gemini 1.5 Pro [5], which offer advanced multimodal reasoning and significantly expanded context windows. This progress is driven by foundational architectural innovations, most notably the Transformer architecture [6], and new training paradigms. This combination presents both significant opportunities and considerable challenges for engineers. For an engineer, choosing the wrong model can lead to inefficient use of resources or system failure.

This challenge extends directly into the educational domain. Future informatics teachers must be prepared to navigate this complex technological landscape, not merely as users, but as specialists who can critically evaluate and select appropriate tools. The development of such professional competencies is a key challenge for modern pedagogy. This creates a need for a unified methodology that can both solve real-world engineering problems and serve as a practical tool for training the next generation of specialists. The main aim of this work is to develop and validate such a unified framework, demonstrating its applicability in both industrial and educational contexts.

2.1. Evolution of LLM Capabilities

The evaluation of LLMs has progressed from static benchmarks to more dynamic, task-oriented assessments. The release of GPT-4 marked a significant milestone, with studies noting its emergent abilities as "sparks of Artificial General Intelligence" [1]. This demonstrated that raw performance on benchmarks was no longer sufficient to capture a model's full potential. Consequently, the research community has shifted towards evaluating LLMs on their ability to perform complex, multi-step tasks, often requiring tool use and planning, which is a core focus of agentic AI research.

2.2. Application of LLMs in Engineering

The engineering domain has become a fertile ground for LLM applications. Recent surveys highlight the growing use of LLMs in Computer-Aided Design (CAD), where they assist in generating design concepts from textual prompts, automating repetitive tasks, and even exploring novel design spaces [7]. Studies have explored the effectiveness of these models in specific fields like mechanical engineering, identifying both significant opportunities and persistent challenges related to factual accuracy and domain-specific reasoning [8, 9]. Beyond design, LLMs are being systematically evaluated for their utility as co-pilots for developers in tasks like code generation, explanation, and debugging [10]. A common theme across recent reviews is that while LLMs show great promise in engineering, their successful integration depends heavily on a structured selection and validation process that accounts for the specific requirements and constraints of each domain [8, 9].

2.3. Pedagogical Alignment and the Research Gap

In parallel, the educational field is grappling with how to best leverage LLMs. Research is emerging on the "pedagogical alignment" of these models—that is, how to fine-tune or guide them to be effective and safe for personalized learning [11]. However, a significant gap exists in the literature: while there is extensive research on using LLMs within engineering and within education, there is a lack of unified frameworks that bridge the two. There is no established methodology that uses the rigorous, quantitative evaluation required for engineering applications as a direct tool for building the corresponding evaluation competencies in future educators. This paper aims to fill this gap by proposing a dual-use framework that is both an engineering tool and a pedagogical model.

3. Materials and Methods

The research was conducted in three main stages:

1. Comparative Analysis of LLM Technical Characteristics. A representative group of models available as of June 2025 was selected. The analysis was conducted by aggregating data from official technical reports and scientific publications.
2. The process, outlined in Figure 1, involves four key steps: (a) establishing key performance indicators (criteria), (b) prioritizing these criteria by assigning weights based on task-specific requirements, (c) standardizing model scores for fair comparison, and (d) calculating a final weighted score to quantitatively determine the best model. The total score for each LLM is calculated as:
   Score = Σ (Weight_i * NormalizedValue_i)
   where Weight_i is the importance of criterion, normalized to a common scale (e.g., 0 to 1).
3. Design of the Unified Framework. The quantitative selection model was established as the "engineering core" of a unified framework, applicable across different domains. For the educational domain, this core was embedded within a "pedagogical shell"—a 4-stage learning cycle.

4. Results

This section describes the unified framework and demonstrates its application in two distinct domains.

4.1. The Engineering Core: Quantitative Selection Process

The engineering core of the framework is a formalized MCDM model, the process of which is visualized in Figure 1. This core provides an objective, data-driven foundation that transforms a potentially suboptimal selection into an optimal, cost-effective integration.

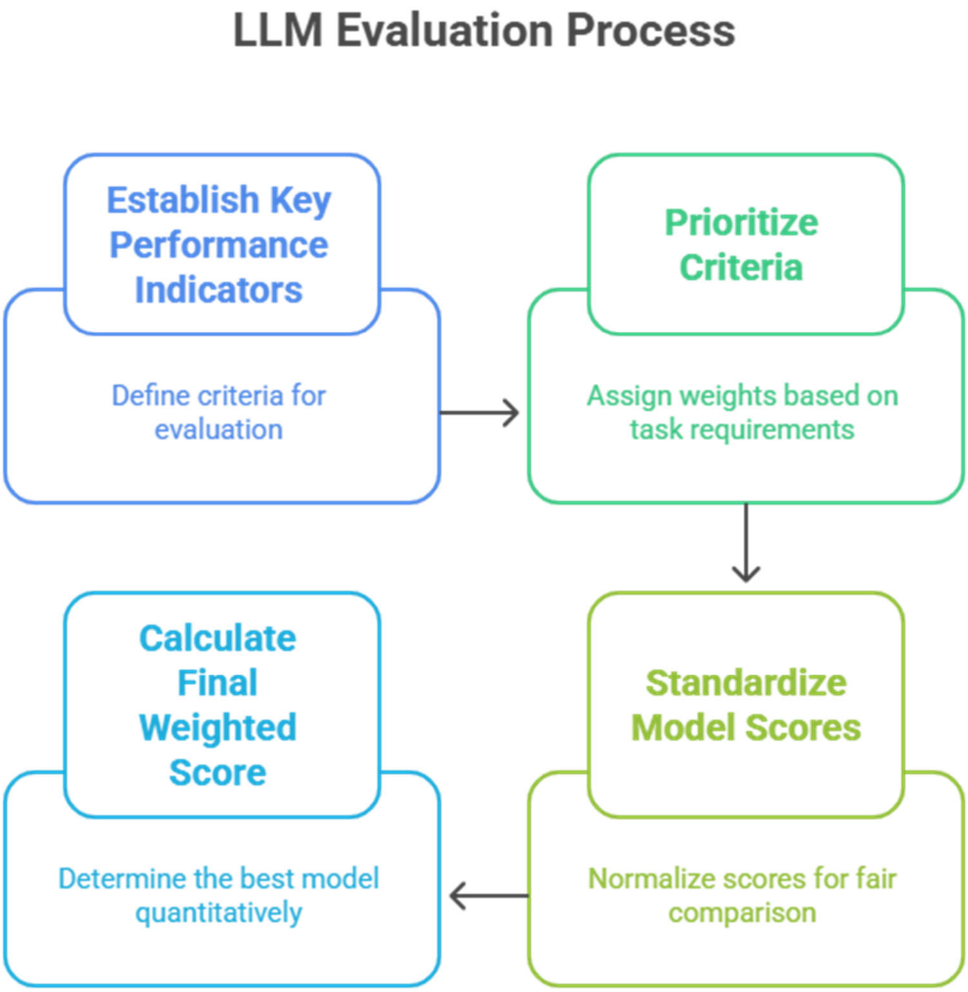

Figure 1. The Engineering Core: Quantitative Selection Process

4.2. Application Domain 1: Industrial Predictive Maintenance

To illustrate the framework in an engineering context, we consider a predictive maintenance system for automated manufacturing. The objective is to integrate an LLM to interpret error logs and sensor data in real-time.

- In this mission-critical scenario, "Privacy & Local Control" (Weight=0.3) and "Reasoning Performance" (Weight=0.3) are paramount.
- Comparative analysis: we apply the quantitative model to compare three candidate models.

Table 1. Quantitative Selection for the Industrial Use Case

Criterion (i)	Weight (W_i)	Claude 3 Opus (Norm. Value)	GPT-4o (Norm. Value)	Llama 3 70B (Norm. Value)
Performance	0.3	0.98	0.95	0.90
Latency	0.2	0.90	0.92	0.80
Privacy	0.3	0.40	0.40	1.00
Cost	0.1	0.60	0.65	0.90
Customizability	0.1	0.50	0.50	0.90
Total Score	1.0	0.70	0.70	0.91

Deployment decision: Meta's Llama 3 is selected due to its superior score, driven by its perfect privacy score and high customizability, which are critical for this use case. This demonstrates the framework's ability to guide high-stakes engineering decisions. This solved industrial case, with its defined criteria and trade-offs, now serves as an authentic, high-fidelity problem-based learning scenario for the pedagogical application domain.

4.3. Application Domain 2: Pedagogical Competency Development

In the educational context, the engineering core is embedded within a 4-stage pedagogical learning cycle. This cycle uses the industrial case study from the previous section as a foundation for a practical, problem-based learning scenario.

Table 2. The 4-Stage Pedagogical Model for Competency Development

Stage of the Model	Corresponding Pedagogical Task for Future Teachers	Developed Professional Competency
1. Analysis	Deconstruct an educational case study and assign weights to selection criteria from the engineering core.	**Analytical & Design Competency:** ability to translate pedagogical needs into a quantitative technical specification.
2. Selection	Use the quantitative model to score and select the optimal LLM. Justify the choice in a technical report.	**Research & Critical Thinking Competency:** ability to conduct evidence-based analysis and make informed decisions.

3. Application	Develop a prototype of the educational tool using the selected LLM.	**Technological & Project Management Competency:** practical skills in AI application, debugging, and iterative development.
4. Reflection	Prepare a presentation analyzing the final prototype's strengths, weaknesses, and the ethical issues it raises.	**Ethical & Communicative Competency:** ability to critically assess the impact of technology and communicate complex ideas.

5. Discussion

The power of the proposed approach lies not in its individual components, but in their integration. The framework is more than just a tool; it is a bridge between industry and education.

5.1. The Holistic Framework in Action

The conceptual model of the framework (Figure 2) illustrates the four critical dimensions that must be balanced. The Quantitative Model (top-left) provides the engineering rigor. This process directly informs Competency Development (bottom-left) by creating practical learning tasks. However, any selection must consider the Ethical Challenges (top-right), such as data governance, and the Systemic Risks (bottom-right), like the exposure of student data. Our unified framework ensures these are not afterthoughts but are integral to the decision process.

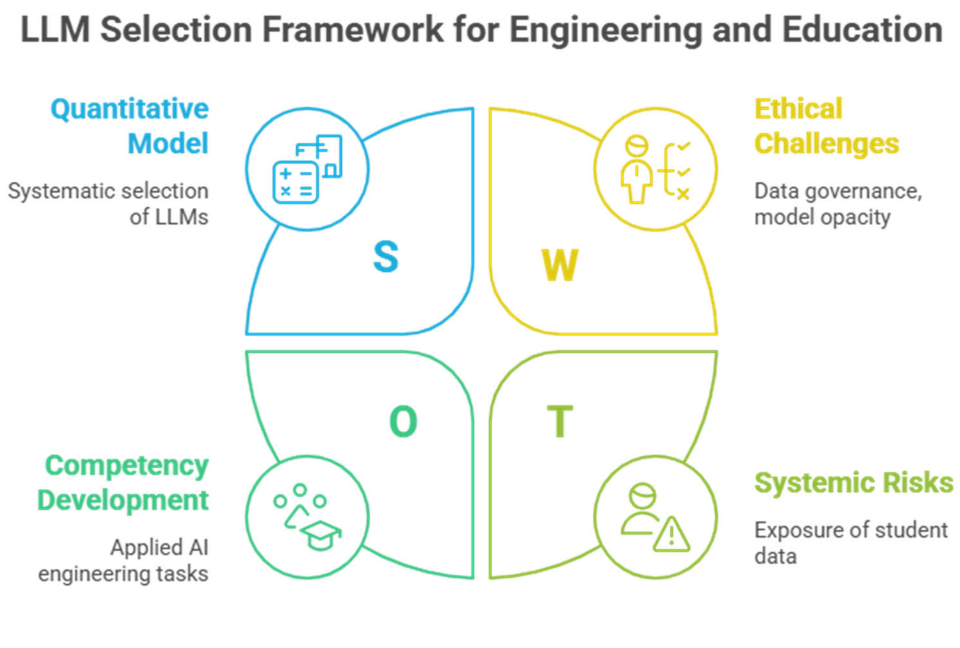

Figure 2. The Conceptual Model of the Unified Framework

5.2. Synergy and Knowledge Transfer Between Industry and Education

The unified framework creates a virtuous cycle of knowledge transfer, as illustrated in Figure 3. Industry-relevant problems and constraints inform the design of authentic learning scenarios for students. Students, trained using the same rigorous, quantitative methods used in industry, graduate with practical, in-demand skills. This closes the gap between academic training and real-world engineering requirements.

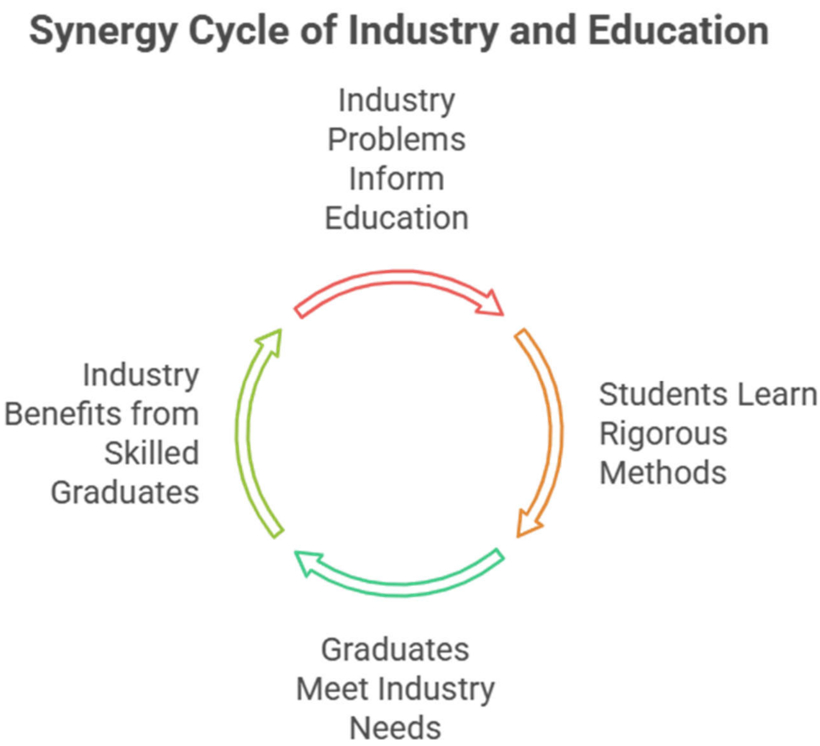

Figure 3. The Synergy Cycle of the Unified Framework

5.3. Implementation as a Didactic Tool

The framework's primary pedagogical strength is its ability to transform an abstract concept (AI evaluation) into a concrete, hands-on activity. By starting with a solved industrial problem, the "cognitive load" on the student is reduced, allowing them to focus on understanding the process of decision-making rather than getting lost in the technical details of an unsolved problem. This "learning by dissecting and recreating" approach is highly effective for complex technical subjects.

5.4. Ethical Implications and Core Trade-Offs

By forcing an explicit weighting of criteria like "Privacy," the framework makes ethical decision-making a core part of the engineering process. This is crucial for future teachers, who will act as gatekeepers of technology in schools. The quantitative model forces a confrontation with the fundamental trade-offs in LLM selection. As visualized in Figure 4, there is often an inverse relationship between a model's raw performance and the level of privacy and control it offers. Proprietary, remote APIs typically lead in performance but require data to be sent to third-party servers. Conversely, local, open-source models offer maximum privacy and customizability, sometimes at the cost of peak performance [12, 13]. The framework provides a structured way to navigate this spectrum.

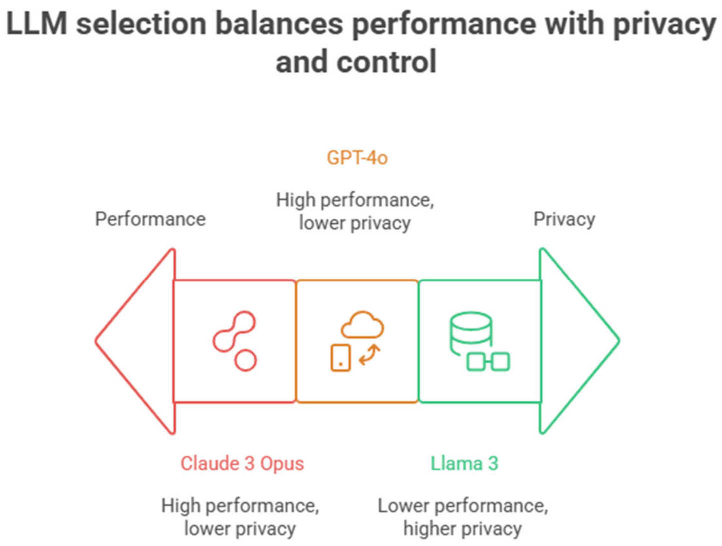

Figure 4. Example of the Performance vs. Privacy Trade-off Spectrum

5.5. Limitations of the Study

This study has several limitations. Firstly, the case study presented in the results section is illustrative; it uses plausible, hypothetical scores for real-world models to demonstrate the framework's application rather than presenting benchmarked performance data. Secondly, the proposed pedagogical model requires empirical validation through longitudinal studies in real teacher training programs.

Conclusions

This paper introduced a unified decision-making framework for the selection and application of LLMs. The key innovation is its demonstrated dual applicability: as a quantitative tool for high-stakes industrial decision-making and as a structured pedagogical model for developing the competencies of future informatics teachers. By creating a synergistic link between industrial needs and educational practice, the framework provides a robust and scalable methodology for preparing a future-ready workforce capable of navigating the complex technical and ethical landscape of artificial intelligence. Future research should focus on the empirical validation of the pedagogical model and the development of an interactive software tool to support the framework's implementation.

Reference

1. Bubeck, S.; Chandrasekaran, V.; Eldan, R.; Gehrke, J.; Horvitz, E.; Kamar, E.; Lee, P.; Lee, Y. T.; Li, Y.; Lundberg, S.; et al. Sparks of Artificial General Intelligence: Early Experiments with GPT-4. arXiv preprint 2023, arXiv:2303.12712. DOI: 10.48550/arXiv.2303.12712.
2. OpenAI. Hello GPT-4o. Available online: https://openai.com/index/hello-gpt-4o/ (accessed on 5 August 2025).
3. Anthropic. Introducing the Next Generation of Claude. Available online: https://www.anthropic.com/news/claude-3-family (accessed on 5 August 2025).
4. Meta AI. Introducing Meta Llama 3: The Most Capable Openly Available LLM to Date. Available online: https://ai.meta.com/blog/meta-llama-3/ (accessed on 5 August 2025).
5. Google. Gemini 1.5 Pro Now Available in 180+ Countries. Available online: https://developers.googleblog.com/en/gemini-15-pro-now-available-in-180-countries-with-native-audio-understanding-system-instructions-json-mode-and-more/ (accessed on 5 August 2025).
6. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; Polosukhin, I. Attention is All You Need. Advances in Neural Information Processing Systems 2017, 30, 5998–6008. DOI: 10.5555/3295222.3295349.
7. Zhang, L.; Le, B.; Akhtar, N.; Lam, S.-K.; Ngo, T. Large Language Models for Computer-Aided Design: A Survey. arXiv preprint 2025, arXiv:2505.08137. Available online: https://arxiv.org/abs/2505.08137 (accessed on 5 August 2025).
8. Gholap, S.; Karande, A. Large Language Model for Requirements Engineering: A Systematic Literature Review. Proceedings of the International Conference on Intelligent Systems and Data Science, India, 2024. Available online:

https://www.researchgate.net/publication/386905816_Large_Language_Model_for_Requirements_Engineering_A_Systematic_Literature_Review (accessed on 5 August 2025).
9. Chiarello, F.; Barandoni, S.; Škec, M. M.; Fantoni, G. Generative Large Language Models in Engineering Design: Opportunities and Challenges. Proceedings of the Design Society 2024, 4, 1959–1968. DOI: 10.1017/pds.2024.198.
10. Huynh, N.; Lin, B. Large Language Models for Code Generation: A Comprehensive Survey of Challenges, Techniques, Evaluation, and Applications. arXiv preprint 2025, arXiv:2503.01245. Available online: https://arxiv.org/abs/2503.01245 (accessed on 5 August 2025).
11. Razafinirina, M.; Dimbisoa, W.; Mahatody, T. Pedagogical Alignment of Large Language Models (LLM) for Personalized Learning: A Survey, Trends and Challenges. Journal of Intelligent Learning Systems and Applications 2024, 16(4), 448–480. DOI: 10.4236/jilsa.2024.164023.
12. Yan, L.; Sha, L.; Zhao, L.; Li, Y.; Martinez-Maldonado, R.; Chen, G.; Jin, Y.; Gašević, D. Practical and Ethical Challenges of Large Language Models in Education: A Systematic Scoping Review. arXiv preprint 2023, arXiv:2303.13379. DOI: 10.48550/arXiv.2303.13379.
13. Mökander, J.; Schuett, J.; Kirk, H. R.; Floridi, L. Auditing Large Language Models: A Three-Layered Approach. AI and Ethics 2024, 4, 1085–1115. DOI: 10.1007/s43681-023-00289-2.