SCIENTIFIC OASIS

Journal of Soft Computing and Decision Analytics

Journal homepage: www.jscda-journal.org
ISSN: 3009-3481

Challenge in Applying ChatGPT in Education: Evaluating Potential Drawbacks through Failure Modes and Effects Analysis and Logarithm Methodology of Additive Weights

Sahand Vahabzadeh[1], Saeid Jafarzadeh Ghoushchi[1], Dragan Pamucar[2,3,4,*]

[1] Faculty of Industrial Engineering, Urmia University of Technology, Urmia, Iran.
[2] UNEC Applied Artificial Intelligence Research Center, Azerbaijan State University of Economics (UNEC), Baku, Azerbaijan
[3] Faculty of Engineering, Dogus University, 34775 Umraniye, Istanbul, Türkiye
[4] School of Engineering and Technology, Sunway University, Selangor, Malaysia

ARTICLE INFO

Article history:
Received 5 January 2026
Received in revised form 7 June 2026
Accepted 26 July 2026
Available online 28 July 2026

Keywords:
Artificial Intelligence; ChatGPT; Failure Modes and Effects Analysis; Logarithm Methodology of Additive Weights; Dominance-Based Nominal Multi-Criteria Analysis.

ABSTRACT

Large language models, such as ChatGPT, are transforming higher education by supporting personalized learning, content generation, and academic assistance. However, their widespread adoption also introduces educational risks that remain insufficiently prioritized in the literature. This study develops an integrated LMAW–DNMA–FMEA framework to identify, prioritize, and evaluate the most critical risks associated with ChatGPT adoption in higher education. Expert judgments are used to assess the relative importance and severity of identified risks, enabling a systematic ranking of potential failure modes. The findings indicate that overreliance on AI-generated content, degradation of critical thinking skills, and unreliable referencing represent the most significant challenges, while issues related to creativity and originality also require attention. Based on the results, targeted mitigation strategies are proposed for educators, developers, and policymakers to support the responsible integration of generative AI into educational environments. The proposed framework provides a structured decision-support approach for AI risk assessment in education and contributes to the development of evidence-based strategies for the effective adoption of large language models in higher education.

1. Introduction

ChatGPT is a groundbreaking technological innovation that is centered upon the application of artificial intelligence (AI) to facilitate machine learning, and is designed to function as a conversationalist using a sophisticated large language model [1]. In addition, the generative pre-trained Transformer (GPT) language model technology-based public tool, ChatGPT, has been developed by OpenAI and has garnered a user base of over one million individuals [2]. ChatGPT has the potential to be a potent information collection tool, given its capacity to accumulate and analyze

* *Corresponding author.*
E-mail address: dpamucar@gmail.com

https://doi.org/10.31181/jscda41202691

© The Author(s) 2026 | Creative Commons Attribution 4.0 International License

extensive quantities of data obtained from a diverse array of sources [3-4]. By offering users access to an extensive range of information and resources, ChatGPT can function as a quasi-virtual library [5].

ChatGPT offers users unrestricted access to a diverse collection of information and resources from around the world, without being bound by temporal or geographical limitations. Its ability to comprehend and respond in multiple languages makes it a valuable tool for users who communicate in different languages. Moreover, ChatGPT is impartial and free from human biases, enabling it to provide users with a more extensive range of resources and recommendations [6]. ChatGPT is a highly advanced chatbot that can perform various text-based tasks, ranging from answering basic queries to completing complex assignments such as generating thank-you notes and solving productivity-related problems. It can even compose entire scholarly essays by dividing the central theme into subtopics and allowing GPT to write each section, resulting in a complete article. The implications of ChatGPT extend beyond its immediate applications, potentially influencing diverse industries [7]. ChatGPT can be a valuable resource in academia, particularly for conducting literature reviews and research, by providing access to a wealth of information and resources [6]. ChatGPT's capacity to supervise the quality of written work could also prove beneficial in the education sector, as it could assist in grading and offering feedback on student assignments [7].

The scientific community and academia have provided mixed reactions to ChatGPT, reflecting the ongoing debate over the advantages versus the drawbacks of advanced AI technologies [8-9]. In order to effectively integrate ChatGPT into the educational sector, it is important to prioritize the benefits and drawbacks associated with its use. By identifying the most significant advantages and disadvantages, educators can develop targeted strategies to address the drawbacks and optimize the benefits of ChatGPT in education. This prioritization process can help educators to make informed decisions regarding the integration of ChatGPT into the learning process, with the ultimate goal of enhancing the academic experience for students.

Yilmaz *et al.*, [10] conducted an analysis of students' perspectives on the use of ChatGPT in programming learning and programming. The research findings indicated that there were both benefits and drawbacks to using ChatGPT in this context. In the year 2023, Sallam [11] conducted a comprehensive and structured examination with the aim of investigating the potential use of ChatGPT in the domains of healthcare education, practice, and research, while also seeking to discern any probable constraints that may arise. Gill [12] research paper delves into the transformative impact of ChatGPT on modern education. The study examines the capabilities of ChatGPT and its potential use in the education sector while also identifying potential concerns and challenges that may arise from its integration [12]. The research aims to provide educators with a comprehensive understanding of ChatGPT's potential to enhance the academic experience for students, as well as the potential pitfalls that should be considered before its implementation. By exploring the benefits and drawbacks of ChatGPT in education, the study provides insights that can aid educators in making informed decisions regarding the use of this advanced AI technology. Adiguzel's [13] research provides a thorough and comprehensive examination of AI technologies, exploring their potential applications in the field of education, as well as the challenges and difficulties associated with their implementation. Bo Zhang's [14] research paper aims to provide an extensive overview of ChatGPT, covering its background, capabilities, potential benefits, and challenges and limitations. The study also explores the impact of technology and AI on educational institutions, highlighting the potential benefits and drawbacks of their integration. By examining the broader implications of ChatGPT and AI in education, the study provides valuable insights that can aid educators in making informed decisions regarding the use of these technologies in the classroom.

The objective of Malinka's [15] research is to evaluate the impact of ChatGPT on higher education, with a specific emphasis on computer security-related specializations. The research endeavors to gather data regarding the efficacy and user-friendliness of ChatGPT in the successful completion of examinations, programming assignments, and term papers. Additionally, the study scrutinizes the potential for misemployment of the tool, encompassing its utilization as an advisor or the replication of its outputs. Although the research underscores the facility with which ChatGPT can be exploited for academic dishonesty, it also deliberates the considerable potential benefits of the tool for the academic arena. Lee [4] focuses on the integration of AI into medical education, highlighting the potential for this technology to transform the way students learn about biomedical sciences. At the same time, the study emphasizes the importance of considering the potential challenges and limitations of ChatGPT, including ethical concerns and potential negative effects [4]. By providing a comprehensive assessment of the integration of ChatGPT into medical education, the study offers valuable insights for educators and researchers seeking to enhance the educational experience of students.

Chukwuere [16] conducted a rapid literature review of recent publications (Jan–Jul 2023) to analyze the use and effectiveness of ChatGPT in higher education, identifying 15 key works. These 15 key works include significant academic articles, conference papers, and book chapters published between January and July 2023 that explore various aspects of ChatGPT's use, advantages, disadvantages, and potential in higher education. Li [17] analyzed Twitter data concerning the use of ChatGPT in education. They employed a sentiment analysis model based on the RoBERTa architecture to categorize the sentiments expressed in the tweets. Their findings highlighted five primary concern categories: academic integrity, effects on learning outcomes and skill development, limitations of ChatGPT's capabilities, policy and social issues, and workforce-related challenges. Farrokhnia [18] employed the SWOT analysis framework to systematically identify and evaluate the strengths and weaknesses of ChatGPT, as well as to explore its potential opportunities and associated threats within the context of education. Montenegro-Rueda [19] in their study systematically reviewed 12 recent studies from leading educational databases, published since ChatGPT's launch in November 2022, to analyze its impact on education. Findings indicate that ChatGPT positively influences the teaching–learning process, but effective implementation depends on adequate teacher training.

1.1. Research Gap and Study Innovation

While existing literature has extensively examined the potential and impact of ChatGPT within educational contexts, a significant research gap remains regarding the systematic evaluation and prioritization of the challenges and disadvantages associated with its adoption. Most studies have focused on its benefits, opportunities, and broad implications; however, they often lack a detailed, structured framework for identifying and ranking the specific risks and limitations that may impede effective implementation. This absence of prioritized assessment hinders stakeholders from understanding which challenges warrant immediate attention and strategic mitigation.

To address this critical gap, the present study introduces a hybrid methodological approach that combines expert elicitation with advanced decision-making techniques. Specifically, the study employs a hybrid MCDM framework to comprehensively evaluate, compare, and prioritize the disadvantages of integrating ChatGPT into educational settings. This approach involves engaging a diverse panel of domain experts, including educators, technologists, and policymakers, to ensure the robustness and relevance of the findings.

The primary innovation of this research lies in its emphasis on identifying the disadvantages and systematically ranking them according to their significance and potential impact. By doing so, the

study provides a nuanced understanding that goes beyond mere enumeration of challenges, offering critical insights into which issues should be addressed as priorities. The resulting insights make a valuable contribution to academic discourse, as well as practical decision-making, assisting educators, educational administrators, policymakers, curriculum developers, educational content creators, researchers, academic community, technology developers, AI providers, in identifying targeted strategies for the responsible and effective deployment of ChatGPT in education. Ultimately, this research aims to foster more informed, strategic, and ethical integration of AI-powered tools within pedagogical practices, promoting sustainable and beneficial outcomes for all stakeholders.

1.2. Research Contribution

This research makes a significant contribution to the growing body of knowledge surrounding AI in education by systematically exploring the complex implications of integrating ChatGPT into learning environments. This study is designed to provide a comprehensive framework for understanding the challenges posed by this innovative technology. In particular, the research aims to assist educators, policymakers, and technologists in making informed decisions about the deployment of ChatGPT in educational settings.

The contributions of this study are as follows:

- Identifying 24 potential disadvantages and challenges of using ChatGPT in education.
- Engaging a diverse range of experts, stakeholders, and DMs to evaluate these challenges.
- Developing a hybrid MCDM approach to assess the disadvantages associated with ChatGPT in education.
- Conducting a comprehensive evaluation and prioritization of the disadvantages of implementing ChatGPT in education to raise awareness among stakeholders.

In group decision-making contexts, the aggregation of expert opinions is critical to ensuring a robust and comprehensive evaluation process. Variations in individual behavior, experience, and inherent uncertainties in decision-making can lead to divergent judgments among experts. Recognizing the importance of these factors, particularly in large-scale studies, it is essential to account for uncertainty and variability in expert assessments to enhance the reliability of the results. In this research, we addressed this issue by incorporating a diverse group of experts from various fields related to decision-making, thereby enriching the breadth of perspectives and minimizing individual bias. Furthermore, additional techniques were employed to strengthen the credibility and reliability of the outcomes, ensuring a more balanced and validated evaluation process.

The research method employed in this study involves the use of three different techniques to evaluate the disadvantages of using ChatGPT in education. The first method used is Failure Modes and Effects Analysis (FMEA) to identify and evaluate potential failures and risks associated with the use of ChatGPT in education. The second method used is the logarithm methodology of additive weights (LMAW) method, a relatively new weighting method, to determine the weight of subjective criteria used to evaluate ChatGPT's effectiveness in education. The third method used is the Dominance-Based Nominal Multi-Criteria Analysis (DNMA) method, another relatively new ranking method, to evaluate the advantages and disadvantages of ChatGPT in education.

The present study's subsequent section is structured as follows: Section 2 expounds upon the employed methodologies, namely FMEA, LMAW, and DNMA. Section 3 provides clarity on the proposed approach. Finally, Section 4 analyzes the implementation of two techniques and reflects on the prioritization of ChatGPT's shortcomings in the realm of education.

2. Methodology

The integration of FMEA, LMAW, and DNMA offers a highly innovative and robust approach that advances traditional MCDM methodologies. FMEA provides a systematic assessment of potential risks and failure modes, while LMAW offers objective, data-driven weighting to accurately capture the relative importance of criteria, reducing subjective biases [20]. DNMA complements this by delivering a non-compensatory, dominance-based ranking that effectively manages conflicting or incomparable alternatives [21]. This combined framework leverages the strengths of each method, resulting in a comprehensive, transparent, and reliable decision-making process that surpasses classic techniques such as AHP, TOPSIS, or Delphi—particularly in complex, multi-faceted evaluations [22]. Methodologically, this integration enhances the rigor and robustness of the analysis also extends its applicability across diverse domains, contributing a versatile, systematic, and credible framework for risk assessment and prioritization beyond its application to ChatGPT's disadvantages in education.

2.1. Failure Modes and Effects Analysis (FMEA)

The FMEA methodology was developed in the 1960s by aerospace industry experts [23], with the primary objective of identifying potential failure modes and assessing the possible causes and outcomes of these failures. By using the Risk Priority Number (RPN) approach, FMEA helps evaluate the risk associated with failure modes and can identify the impact of such failures on the management process [24]. FMEA is a highly effective methodology that can identify and eliminate potential or known failures in complex systems, thereby enhancing their safety and reliability [25]. The assessment of each failure mode is based on three parameters: difficulty of detection (D), likelihood of occurrence (O), and severity (S), which are evaluated using a numerical scale of 1 to 10. The RPN is calculated by multiplying the probabilities of the severity, occurrence, and detection of the failure mode, with higher RPN values indicating greater risk. Table 1 in the article presents a 10-point scale for assessing the risk factors of severity, likelihood of occurrence, and difficulty of detection [26]. This approach helps in determining the risk priorities of different failure modes and can guide informed risk management decisions. The presence of replicated numerical values within the RPN components is evident. Occasionally, multiple risk factors may exhibit identical RPN scores. It is worth noting that the conventional FMEA methodology lacks comprehensive techniques for prioritizing risk evaluations and often perplexes risk management stakeholders [27].

In Hybrid FMEA, measuring the range (R) or domain of a risk refers to determining the extent or boundaries of the potential impact of the risk. This involves identifying the range of values or conditions in which the risk could occur and the potential consequences of that occurrence [28]. Measuring the range of a risk is important because it helps to provide a more accurate understanding of the potential impact of the risk on the system or process being analyzed. By understanding the range of the risk, organizations can better understand the conditions under which the risk is likely to occur and the potential consequences of that occurrence [29]. This information can help organizations to develop more effective risk management strategies and controls to mitigate the risk. To measure the range of a risk, organizations can use various techniques, such as sensitivity analysis or modeling, to identify the potential values or conditions where the risk is most likely to occur. By analyzing the data and identifying these ranges, organizations can gain a better understanding of the potential impact of the risk and develop more effective risk management strategies. In summary, measuring the range or domain of a risk in hybrid FMEA is important because it helps to provide a more accurate understanding of the potential impact of the risk. By understanding the range of the

risk, organizations can develop more effective risk management strategies and controls to mitigate the risk.

$$RPN = S \times O \times D \times R \quad (1)$$

Table 1
Hybrid ratings for Severity, Occurrence, Detection, and Range factors

Rating	Severity (S)	Occurrence (O)	Detection (D)	Range(R)
10	Hazardous without warning	Very high: failure is almost inevitable	Absolute uncertainty	Unprecedentedly high: The most extensive scope of impact
9	Hazardous with warning			
8	Very high	High: repeated failures	High: repeated failures	Very high
7	High			
6	Moderate	Moderate: occasional failures	Moderate: occasional failures	Moderate
5	Low			
4	Very low			
3	Minor	Low: relatively few failures	Low: relatively few failures	Low
2	Very minor			
1	None	Remote: failure is unlikely	Remote: failure is unlikely	Very low

2.2. Logarithm Methodology of Additive Weights (LMAW) method

The LMAW technique is a recent method in the field of criterion weighting and alternative ranking, introduced by [30]. This technique involves determining weight coefficients for different criteria and is presented in detail in the academic article. The implementation procedure for the LMAW approach is described in several steps, as outlined by [31] within this section. The weight coefficients for the criteria are provided, and the process of applying this technique is well-defined, enabling users to follow the steps in a clear and straightforward manner. Overall, the LMAW technique is a useful tool for evaluating and ranking alternatives based on multiple criteria, and its introduction in the academic article provides readers with a valuable methodological contribution [20].

A set of alternatives, denoted as *B*, which includes $\{B_1, B_2, \dots, B_n\}$. The evaluation is conducted based on a set of *m* criteria, denoted as *D*, which consists of $\{D_1, D_2, \dots, D_m\}$. The process involves the participation of *k* specialists, denoted as *P*, which includes $\{P_1, P_2, \dots, P_k\}$. The evaluation is conducted using a pre-established linguistic scale, which enables the specialists to assess and contrast the alternatives based on the evaluation criteria.

Step 1: The outlines a process for normalizing the decision matrix that involves prioritizing criteria $D = \{D_1, D_2, \dots, D_m\}$ based on a linguistic scale previously defined by a cluster of specialists denoted as $P = \{P_1, P_2, \dots, P_k\}$. The prioritization exercise entails assigning higher values to more important criteria and lower values to less important criteria on the linguistic scale. This process results in the generation of a priority vector $F^e = \{\gamma_{D_1}^e, \gamma_{D_2}^e, \dots, \gamma_{D_m}^e\}$, where $\gamma_{D_m}^e$ represents the linguistic scale value assigned by specialists $e(1 \le e \le k)$ to criterion $D_t(1 \le t \le m)$. In other words, the priority vector reflects the priority given to each criterion by the specialists' cluster, based on their collective assessment of the criteria's relative importance. This process of normalizing the decision matrix is an essential step in multi-criteria decision-making, as it helps to standardize the evaluation process and ensure that criteria are considered fairly and objectively. Overall, the priority vector generated through the process of normalizing the decision matrix provides a valuable tool for DMs, enabling them to make informed decisions based on expert assessments of criterion importance.

Step 2: Defining the absolute anti-ideal point, denoted as γ_{AIP}. This point is determined as being lower than the minimum value found in the priority vector. The calculation of γ_{AIP} is based on the following equation $\gamma_{AIP} = \frac{\gamma^e_{min}}{S}$, where γ^e_{min} represents the minimum value of the priority vector, and *s* is a value greater than the base of the logarithmic function. If the *Ln* function is used, it is recommended to select the value of *s* as 3.

Step 3: Determining the correlation between the priority vector and the absolute anti-ideal point. This correlation is calculated using Equation (2), which involves dividing the linguistic scale value assigned to a criterion by the absolute anti-ideal point:

$$\eta^e_{D_m} = \frac{\gamma^e_{D_m}}{\gamma_{AIP}} \tag{2}$$

As a result, the relation vector $R^e = (\eta^e_{D_1}, \eta^e_{D_2}, \dots, \eta^e_{D_m})$ is derived. The parameter $\eta^e_{D_m}$ represents the value obtained from the relation vector based on Equation (2) and is used to generate the relation vector of $e(1 \le e \le k)$.

Step 4: determining the vector of weight coefficients $w_j = (w_1, w_2, \dots, w_m)^T$ for specialists' evaluation. This involves computing the weight coefficients for specialists $\mathrm{e}(1 \le \mathrm{e} \le \mathrm{k})$ using Equation (2). These weight coefficients, denoted as

$$w^e_j = \frac{log_A(\eta^e_{D_m})}{log_A(\prod^m_{j=1} \eta^e_{D_m})}, B > 1 \tag{3}$$

are derived from the evaluations of the e^{th} specialists', with $\eta^e_{D_m}$ representing the elements of the relation vector *R*. The weight coefficients must satisfy the condition of $\sum^m_{j=1} w^e_j = 1$. To obtain the aggregated vector of weight coefficients, the Bonferroni aggregator is applied using Equation (4), where:

$$w_j = (\frac{1}{k.(k-1)} . \sum^k_{x=1} (w_j{}^{(x)})^f . \sum^k_{y=1, y\neq x} (w_{ij}{}^{(y)})^q)^{\frac{1}{p+q}} \tag{4}$$

The stabilization parameters *p* and *q* are used in Equation (4), with *p* and *q* being greater than or equal to zero $(p, q \ge 0)$. It is important to note that the resulting weight coefficients should satisfy the condition of $\sum^m_{j=1} w_j = 1$.

2.3. Dominance-Based Nominal Multi-Criteria Analysis (DNMA) method

In 2019, Liao and Wu introduced the DNMA method as a new approach for listing alternatives. This method employs two different normalized techniques, namely linear and vector, and three different joining functions, namely full compensation-CCM, incomplete compensation-UCM, and incomplete compensator-ICM. The implementation steps of this method are outlined in the article by Liao [32]. In summary, the DNMA method represents a novel approach to multi-criteria decision-making, leveraging a range of techniques and functions to help DMs evaluate and compare alternatives based on multiple criteria.

Step 1 of the method involves normalizing the decision matrix. Linear normalization $(\tilde{\mathrm{x}}^{1\mathrm{N}}_{\mathrm{ij}})$ is achieved using Equation (5), while vector normalization $(\tilde{\mathrm{x}}^{2\mathrm{N}}_{\mathrm{ij}})$ is achieved using Equation (6). Linear normalization is calculated as

$$\tilde{x}^{1N}_{ij} = 1 - \frac{|x^{ij} - r_j|}{\max\{\max_i x^{ij}, r_j\} - \min\{\min_i x^{ij}, r_j\}} \tag{5}$$

while vector normalization is calculated as

$$\tilde{x}^{2N}_{ij} = 1 - \frac{|x^{ij} - r_j|}{\sqrt{\sum^m_{i=1} (x^{ij})^2 + (r_j)^2}} \tag{6}$$

The r_j value represents the target value for the D_j criterion and is considered as $\max_i x^{ij}$ for benefit criteria and $\min_i x^{ij}$, r_j for cost criteria.

Step 2 of the DNMA method, as described in the academic article, involves determining criterion weights through a three-phase process.

In the first phase, the standard deviation (σ_j) of the criterion D_j is calculated using Equation (7), where *n* is the number of alternatives. The standard deviation is computed using the formula:

$$\sigma_j = \sqrt{\frac{\sum_{i=1}^{n}\left(\frac{x^{ij}}{\max_i ij} - \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x^{ij}}{\max_i ij}\right)\right)^2}{n}} \tag{7}$$

Next, the standard deviation values are normalized using Equation (8), which is expressed as

$$w_j^{\sigma} = \frac{\sigma_j}{\sum_{i=1}^{m}\sigma_j} \tag{8}$$

Finally, the weights are adjusted by applying Equation (9), which is given as

$$\widetilde{w}_j = \frac{\sqrt{w_j^{\sigma}.w_j}}{\sum_{i=1}^{m}\sqrt{w_j^{\sigma}.w_j}} \tag{9}$$

This process ensures that the weights are adjusted based on the standard deviation of the criteria, resulting in a more accurate and reliable assessment of the alternatives.

Step 3 of the DNMA method, the aggregation models are calculated for each alternative using three separate aggregation functions: CCM, UCM, and ICM. The complete compensatory model (CCM) is determined using Equation (10), where:

$$u_1(a_i) = \sum_{j=1}^{m}\frac{\widetilde{w}_j.x_{ij}^{1M}}{\max_i \tilde{x}_{ij}^{1M}} \tag{10}$$

The uncompensatory model (UCM) is calculated using Equation (11), where:

$$u_2(a_i) = \max_j \widetilde{w}_j\left(\frac{1-x_{ij}^{1M}}{\max_i \tilde{x}_{ij}^{1M}}\right) \tag{11}$$

The incomplete compensatory model (ICM) is evaluated using Equation (12), where:

$$u_3(a_i) = \prod_{j=1}^{m}\left(\frac{x_{ij}^{2M}}{\max_i \tilde{x}_{ij}^{2M}}\right)^{\widetilde{w}_j} \tag{12}$$

These equations allow for the calculation of the three different aggregation models, each employing a distinct formula for determining the overall score of each alternative. The CCM and UCM models are compensatory in nature, while the ICM model is non-compensatory.

Step 4 of the DNMA method involves the critical task of integrating the utility values obtained in Step 3. This is accomplished using Equation (13), which operates on the Euclidean distance principle. Specifically, DM_i is determined by the expression:

$$\begin{aligned} DM_i = {} & w_1\sqrt{\varphi\left(\frac{u_1(a_i)}{\max_i u_1(a_i)}\right)^2 + (1-\varphi)\left(\frac{n-r_1(a_i)+1}{n}\right)^2} \\ & -w_2\sqrt{\varphi\left(\frac{u_2(a_i)}{\max_i u_2(a_i)}\right)^2 + (1-\varphi)\left(\frac{r_2(a_i)}{n}\right)^2} \\ & +w_3\sqrt{\left(\frac{u_3(a_i)}{\max_i u_3(a_i)}\right)^2 + (1-\varphi)\left(\frac{n-r_3(a_i)+1}{n}\right)^2} \end{aligned} \tag{13}$$

This equation combines the weights from Step 2 and the utility values from Step 3 to generate a single value (DM_i) that represents the overall desirability of each alternative. This process is essential

in ensuring effective integration of the various utility functions and provides DMs with a clear and concise way to compare the alternatives.

The aforementioned equation employs $r_1(a_i)$ and $r_3(a_i)$ as indicators of the sequence number for alternative a_i, which is ranked in descending order based on the CCM and ICM functions, respectively. The sequence number for the UCM utility function is represented by $r_2(a_i)$, arranged in ascending order with the smallest value first. The φ coefficient serves as a relative measure of the subordinate utility values, ranging between 0 and 1, with a recommended value of $\varphi = 0.5$. Additionally, the coefficients w_1, w_2 and w_3 correspond to the importance weights of the CCM, UCM, and ICM utility functions, respectively. These weights are assigned by the decision maker, with the sum of weights being equal to 1 ($w_1 + w_2 + w_3 = 1$). When assigning the weights, the decision maker may prioritize a comprehensive performance evaluation of the alternatives and hence assign the greatest weight to w_1. If the decision maker is risk-averse and desires that the selected alternative does not perform poorly according to certain criteria, they may assign the greatest weight to w_2. Alternatively, if the decision maker wishes to consider both the overall performance and the risks, they may assign the greatest weight to w_3. Finally, the *DN* values are sorted in descending order, and the alternative with the highest value is regarded as the most desirable. Figure 1 summarizes the processing steps of the DNMA method.

3. Proposed Approach

The FMEA approach entails the identification of prospective failure modes, their corresponding effects, and the probability of their occurrence. This methodology aids in the recognition of potential risks associated with the utilization of ChatGPT within the educational domain.

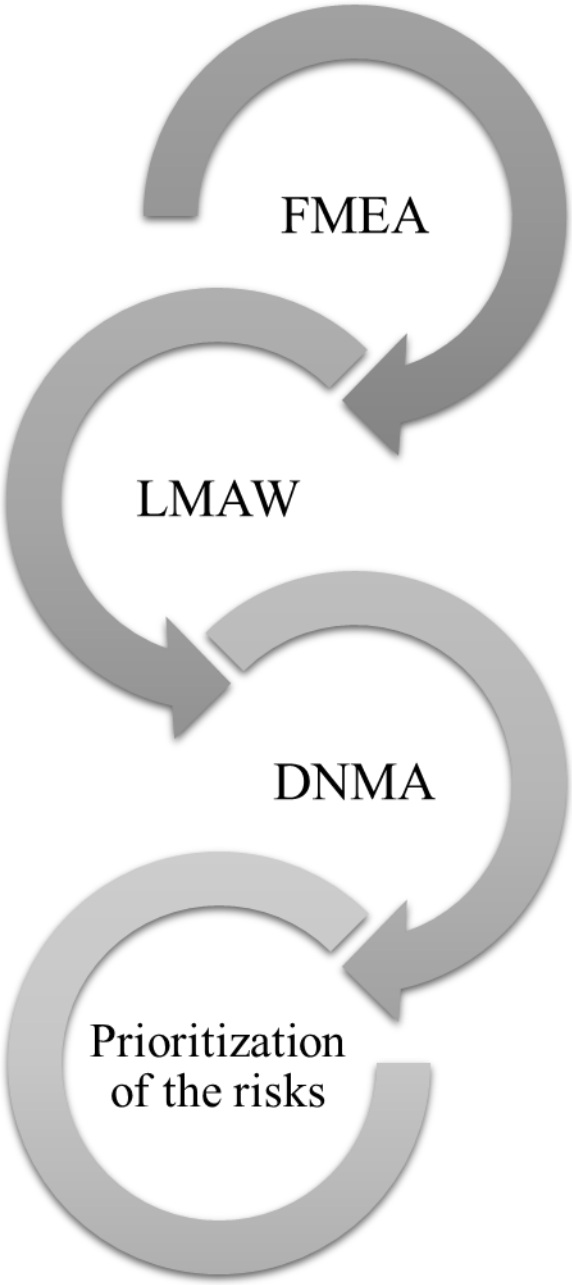
The LMAW technique is employed to ascertain the magnitude of subjective criteria employed in the assessment of ChatGPT's efficacy within the educational sphere. This approach encompasses the assignment of weights to diverse criteria based on their relative significance, subsequently employing these weights to rank and evaluate various alternatives.

The DNMA method is employed to assess the merits and demerits of incorporating ChatGPT into education. This method entails ordering different alternatives according to their superiority, inferiority, or non-comparability with respect to specific criteria.

In general, the research methodology employed in this investigation incorporates a blend of qualitative and quantitative methodologies to evaluate the advantages and disadvantages of employing ChatGPT within the educational realm. The utilization of multiple techniques serves to ensure the robustness and dependability of the findings.

3.1. Definition of the Problem

This research employs a combination of qualitative and quantitative approaches to rank the drawbacks of incorporating ChatGPT into educational contexts. The investigation utilizes multi-criteria evaluation techniques and aims to solicit expert viewpoints in order to evaluate various factors that might potentially influence the efficacy of ChatGPT in education. By means of expert opinions, the study endeavors to provide a comprehensive understanding of the limitations associated with the utilization of ChatGPT in education, as well as to identify the most crucial factors that necessitate consideration when implementing such technology. On the whole, this research furnishes a comprehensive assessment of the disadvantages of employing ChatGPT in education, while considering the perspectives of field experts. The study's outcomes can serve as a valuable asset for educators and institutions contemplating the integration of AI-based chatbots in educational environments.



- Identifying the potential failure modes according to the risk assessment scope
- Determining the SODR factors for each failure by the FMEA team
- Determining the reliability of each failure by the FMEA team

- Determination of criterion weight coefficients
- Calculation anti-ideal points
- Determination of relationship vector
- Calculation of weight coefficient vector

- Normalization of the decision matrix
- Adjustment of criterion weights
- Calculation of utility functions
- Combining utility functions

Fig 1. Chart of the Proposed Model

3.2. Explanation of the Disadvantages

The article presents a comprehensive evaluation of the disadvantages of using ChatGPT in educational settings. The authors gathered opinions and suggestions from experts in the field to identify the drawbacks of utilizing ChatGPT in education. The experts listed several potential disadvantages, such as ethical concerns, technical limitations, and the potential for the technology to promote laziness or dependence on technology, Table 2. The authors stress the importance of considering these factors when implementing AI-based chatbots in educational settings.

Table 2
Opinions on disadvantages of using ChatGPT in the Education

Symbol	Challenges
D1	The use of ChatGPT can lead individuals to become lazier or rely too heavily on the technology.
D2	ChatGPT can cause anxiety related to education or learning.
D3	There is a possibility that ChatGPT may not always provide accurate answers to questions.
D4	Limited information available can hinder the ability of ChatGPT to answer some questions.
D5	The answers provided by ChatGPT may not always meet the desired expectations.
D6	The lack of a real training environment or tool can limit the capabilities of ChatGPT.
D7	ChatGPT may have a negative impact on the development of critical thinking skills.
D8	The use of ChatGPT may lead to increased self-dependence and reduced human interaction.
D9	Ethical issues, such as bias, plagiarism, and data privacy and security concerns, can arise with the use of ChatGPT.
D10	There is a risk of incorrect or inaccurate information being provided by ChatGPT.
D11	Transparency issues can arise with the use of ChatGPT.
D12	Inaccurate or inadequate referencing/citation can be a problem with ChatGPT content.
D13	Legal issues can arise with the use of ChatGPT.
D14	ChatGPT may have limited knowledge from before 2021.
D15	ChatGPT content may contain unnecessary or excessive detail.
D16	Lack of originality can be a problem with ChatGPT content.
D17	Copyright issues can arise with the use of ChatGPT.
D18	ChatGPT may have limited understanding of language and context.
D19	ChatGPT may struggle with handling complex tasks or issues.
D20	The quality and diversity of training data can affect the performance of ChatGPT.
D21	ChatGPT may lack creativity and originality in their responses.
D22	Difficulty understanding accents, dialects or non-standard language use can be a problem for ChatGPT.
D23	ChatGPT may rely on pre-programmed responses and decision trees, limiting their flexibility and adaptability.
D24	Technical errors or glitches can disrupt or invalidate conversations with ChatGPT.

4. Methods Application

The objective of this section is to illustrate the application of the proposed methodology.

4.1. FMEA Application

In FMEA, the risk associated with potential failure modes is assessed and quantified using a Risk Priority Number (RPN) calculation, which is determined by multiplying Severity, Occurrence, Detection and Range scores obtained from Table 3.

Table 3
Identified disadvantages of using ChatGPT in Education

Symbol	Severity (S)			Occurrence (O)			Detection (D)			Range (R)		
	TM1	TM2	TM3	TM1	TM2	TM3	TM1	TM2	TM3	TM1	TM2	TM3
D1	8	7	9	9	8	8	3	6	4	8	7	8
D2	5	4	4	3	4	4	8	8	9	4	4	5
D3	5	4	5	4	2	3	2	3	1	2	1	2
D4	3	4	4	6	7	6	3	4	2	3	2	2
D5	6	7	6	5	4	5	2	4	2	4	3	4
D6	8	6	7	6	5	5	5	4	6	5	4	4
D7	7	8	8	7	4	6	7	8	7	6	6	5

Table 3
Continued

Symbol	Severity (S)			Occurrence (O)			Detection (D)			Range (R)		
	TM1	TM2	TM3	TM1	TM2	TM3	TM1	TM2	TM3	TM1	TM2	TM3
D8	6	7	6	8	9	9	3	5	4	3	4	3
D9	5	7	6	6	8	6	5	6	4	4	5	5
D10	5	4	6	3	2	4	3	2	2	2	1	2
D11	7	6	7	6	7	9	9	8	8	3	3	2
D12	8	8	7	8	6	6	8	6	7	5	4	6
D13	9	8	8	3	5	4	7	6	6	3	2	3
D14	6	5	7	2	3	3	8	8	9	2	2	1
D15	5	6	6	6	4	6	2	2	1	6	3	4
D16	7	6	7	6	7	5	6	5	5	3	4	2
D17	8	9	8	4	5	3	5	4	5	2	3	3
D18	7	6	8	3	4	3	4	3	4	4	2	2
D19	8	9	9	5	7	6	2	3	1	8	7	7
D20	7	8	8	6	5	5	7	8	7	5	3	4
D21	7	7	6	8	6	7	6	5	6	7	5	6
D22	6	7	7	2	4	3	4	5	4	6	4	4
D23	9	7	8	6	6	4	6	4	5	5	5	6
D24	8	6	8	8	7	8	3	2	2	6	5	7

4.2. LMAW Application

The process of calculating the weight coefficients of the criteria involved the prioritization of the criteria by three experts, who utilized the scale provided in Table 4.

Table 4
Prioritization Scale [30]

Linguistic Variables	Prioritization Score
Absolutely Low (AL)	1
Very Low (VL)	1.5
Low (L)	2
Medium (M)	2
Equal (E)	3
Medium High (MH)	3.5
High (H)	4
Very High (VH)	4.5
Absolutely High	5

Due to the involvement of three professionals in the evaluation process, three priority vectors were discerned. These priority vectors represent the relative importance of the different criteria and were determined based on the experts' rankings and assessments. The use of multiple experts in the assessment process can help to ensure that the identified priorities are based on a diverse range of perspectives and expertise.

$$E^1 = \{\text{H, VH, MH, AH}\}$$
$$E^2 = \{\text{MH, H, E, H}\}$$
$$E^3 = \{\text{H, AH, MH, VH}\}$$

In order to establish the relationship between the elements of the priority vector and the absolute anti-ideal point (γ_{AIP}), the value of γ_{AIP} was arbitrarily defined as 0.5. Using Equation (2) and the data gathered from the expert priority vectors and the γ_{AIP} value, the specific relationships

between the priority vector elements and the γ_{AIP} were determined. The subsequent section of the analysis presents these relationships in greater detail.

The components of the vector R^1 were derived using Equation (4), which was applied in the following manner.

$$\eta_{K1}^1=8, \eta_{K2}^1 = 9, \eta_{K3}^1 = 7, \eta_{K4}^1=10$$
$$R^1 = (8, 9, 7,10)$$

Similarly, the components of the vectors R^2 and R^3 were obtained using the same method as employed for the vector R^1.

$$R^2 = (7, 8, 6,8)$$
$$R^3 = (8, 10, 7, 9)$$

To determine the weight coefficients vector, the components of the vector w_j^1 for the first expert were calculated using Equation (3), which was applied in the following manner.

$$w_1^1 = \frac{\ln(8)}{\ln(8\times9\times7\times10)}=0.244$$
$$w_2^1 = \frac{\ln(9)}{\ln(8\times9\times7\times10)}=0.258$$
$$w_3^1 = \frac{\ln(7)}{\ln(8\times9\times7\times10)}=0.228$$
$$w_4^1 = \frac{\ln(10)}{\ln(8\times9\times7\times10)}=0.270$$
$$w_j^1 = (0.244; 0.258; 0.228; 0.270)$$

The weight coefficient values obtained from this calculation satisfy the condition that the sum of all w_j^1 coefficients for j=1 to 4 equals 1 ($\sum_{j=1}^4 w_j^1 = 1$). The remaining elements of the vectors w_j^2 and w_j^3 were calculated in a similar manner to that of w_j^1.

$$w_j^2 = (0.246, 0.263, 0.227, 0.263)$$
$$w_j^3 = (0.244, 0.270, 0.228, 0.258)$$

By utilizing Equation (4), it is possible to derive the weighting vector aggregate of coefficients.

$$w_j = (0.2448, 0.2637, 0.2278, 0.2637)^T$$

To illustrate, the value of $w_1 = 0.2448$ was computed by taking the average of the values of $w_j^e (1 \leq e \leq 3)$ for experts, where $w_1^1 = 0.244$, $w_1^2 = 0.246$ and $w_1^3 = 0.244$.

Similarly, the remaining values of the weight coefficient vector were calculated using the same method as employed for the value of w_1.

4.3. DNMA Application

A panel of subject matter experts from various disciplines were consulted to develop the decision matrix assessing potential disadvantages. Each expert was tasked with evaluating each alternative according to a set of predetermined criteria in a rubric format. Specifically, if an expert believed a disadvantage could reasonably occur based on their area of expertise and the criteria, they would assign one point to that alternative-criterion combination. However, if they determined through their evidence-based analysis that a particular disadvantage outlined in the criteria was implausible or unlikely to come to fruition based on current research and literature, no points would be allocated for that alternative under review. This methodology allowed for a systematic and scholarly appraisal of alternatives that incorporated diverse perspectives and specialized knowledge to reasonably

identify most probable disadvantages, thereby strengthening the decision matrix developed to inform further evaluation. The results of the experts' assessments were then aggregated and documented in Table 5 to provide a clear and transparent means for reviewing the compiled decision matrix, thereby strengthening subsequent analysis and consideration of alternatives by incorporating diverse perspectives from the scholarly community.

Table 5
The Decision matrix of the Disadvantages of using ChatGPT in Education

Symbol	S	O	D	R
D1	24	25	13	23
D2	13	11	25	13
D3	14	9	6	5
D4	11	19	9	7
D5	19	14	8	11
D6	21	16	15	13
D7	23	17	22	17
D8	19	26	12	10
D9	18	20	15	14
D10	15	9	7	5
D11	20	22	25	8
D12	23	20	21	15
D13	25	12	19	8
D14	18	8	25	5
D15	17	16	5	13
D16	20	18	16	9
D17	25	12	14	8
D18	21	10	11	8
D19	26	18	6	22
D20	23	16	22	12
D21	20	21	17	18
D22	20	9	13	14
D23	24	16	15	16
D24	22	23	7	18

Appendix, Table A1 presents the outcomes of linear normalization, which were obtained by utilizing Equation (5) on the data presented in Table 5.

Appendix, Table A2 shows the vector normalization outcomes, which were obtained by employing Equation (6) on the data presented in Table 5.

In order to modify the criterion weights, the standard deviations of the criteria were initially computed using Equation (7) and are presented in Appendix, Table A3.

The standard deviation values were normalized using Equation (8), and the normalized values are presented in Appendix, Table A4.

The modified weight values were determined using Equation (9), and the resulting values are shown in Appendix, Table A5.

Appendix, Table A6 lists the values calculated using Equations (10), (11), and (12) for the computation of the utility functions CCM, UCM, and ICM, respectively.

Experts have determined that the appropriate values for the parameters φ, w_1, w_2 and w_3 are 0.5, 0.3, 0.1, and 0.6, respectively. Using these values and Equation (13), the performance scores for the different alternatives were calculated and are presented in Appendix, Table A6, which also shows

the rankings of the alternatives based on these scores. The calculation of these scores was based on the Euclidean distance method.

5. Results

The present study analyzed and compared the results of two different methods, LMAW-DNMA and Hybrid FMEA, for prioritizing the use of ChatGPT in education. The results of this analysis are presented in Table 6 and Figure 2.

Table 6
Comparison of disadvantages of using ChatGPT in education based on two approaches

Symbol	Hybrid FMEA		LMAW-DNMA	
	RPN	Rank	Utility Values	Rank
D1	2214.815	1	0.845466	5
D2	573.7654	14	0.583186	21
D3	46.66667	24	0.590402	18
D4	162.5556	22	0.589287	19
D5	288.9877	18	0.662844	15
D6	808.8889	9	0.7654	10
D7	1805.358	2	0.872255	2
D8	731.8519	12	0.605346	16
D9	933.3333	8	0.862583	4
D10	58.33333	23	0.588022	20
D11	1086.42	7	0.711232	12
D12	1788.889	3	0.865934	3
D13	562.963	15	0.60186	17
D14	222.2222	20	0.555183	24
D15	218.2716	21	0.796351	7
D16	640	13	0.791645	9
D17	414.8148	16	0.567155	22
D18	228.1481	19	0.560362	23
D19	762.6667	11	0.690754	14
D20	1199.407	5	0.793418	8
D21	1586.667	4	0.912324	1
D22	404.4444	17	0.718936	11
D23	1137.778	6	0.707061	13
D24	787.1111	10	0.806608	6

The LMAW-DNMA methodology unveiled that the drawbacks linked to D21 and D7 were ranked as the foremost and second highest precedence, correspondingly. Furthermore, D12 was positioned third, while D14 was positioned as the least significant in terms of priority. These findings demonstrated consistency with the outcomes derived from the Hybrid FMEA approach, which yielded a very analogous prioritization of the drawbacks. Specifically, the FMEA methodology designated D1 as the highest priority (RPN=2214), succeeded by D7 in the second rank (RPN=1805), and D12 in the third rank (RPN=1788). Conversely, D3 was identified as the drawback with the least priority coefficient, with an RPN value of 46.

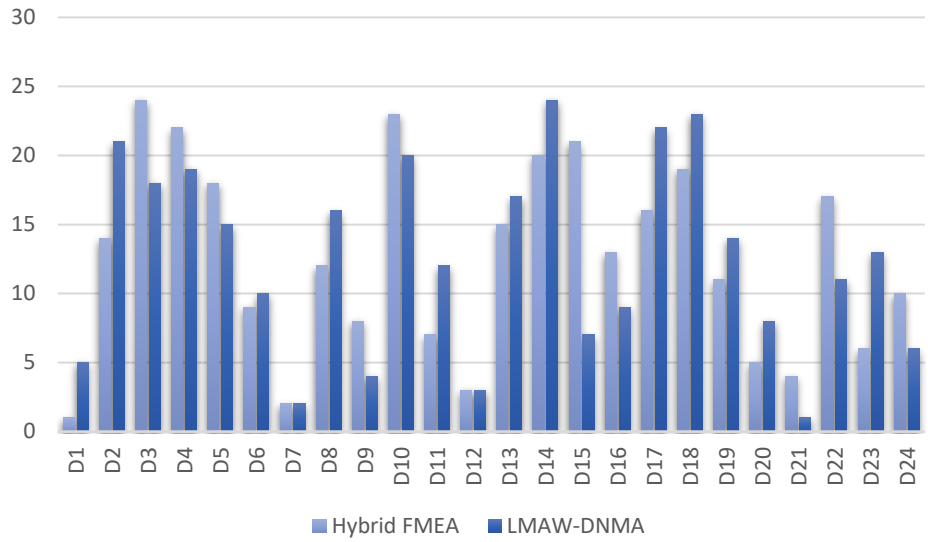


Fig 2. A Bar Chart Analysis of Two Approaches

5.1. Validation of Proposed Approach

After the initial results were obtained, it was important to check if these findings were reliable, accurate, and effective. Therefore, an essential step in MCDM involves testing the stability of the results by comparing them with those from different decision-making methods [33].

To preliminarily substantiate the reliability and precision of the proposed framework and its outputs, a meticulous sensitivity analysis will be applied to the method's criteria. As a critical element of simulation-based experimentation—given its capacity to affect model construction—this analysis is conventionally deployed to investigate system behavior through scenarios that feature modulated criterion weighting.

The Second validation study involved comparing the results from the proposed method (LMAW-DNMA) with two existing methods, namely LMAW-TODIM [34] and LMAW-WASPAS [35]. This comparison aimed to evaluate how well the proposed method worked in identifying and prioritizing the challenges in implementing ChatGPT in education. In the next validation step, Pearson's correlation analysis was used to verify the consistency of the results obtained from the MCDM techniques.

5.1.1. Sensitivity analysis

The primary objective of this analysis was to scrutinize the effect of variable criterion weighting on the ultimate ranking order. To this end, seven discrete scenarios were methodically generated through systematic manipulations of the criterion weights, with each configuration representing a distinct weighting scheme to facilitate a thorough evaluation of the method's parametric sensitivity. A comparative assessment of the resultant rankings was subsequently conducted to detect variations, and the examination of these fluctuating weight values afforded a comprehensive interpretation of the experimental dynamics. These results, along with the corresponding ranking fluctuations, are systematically illustrated in Figure 3, which visually elucidates the influence of divergent weighting paradigms on the prioritization of challenges associated with integrating ChatGPT into education. Beyond merely validating the robustness and reliability of the proposed framework, this sensitivity analysis critically underscores the pivotal significance of weight assignments and the subjective input of expert judgment. Ultimately, the empirical findings firmly

establish that any adjustment to the criterion weighting parameters induces notable deviations in the eventual outcomes.

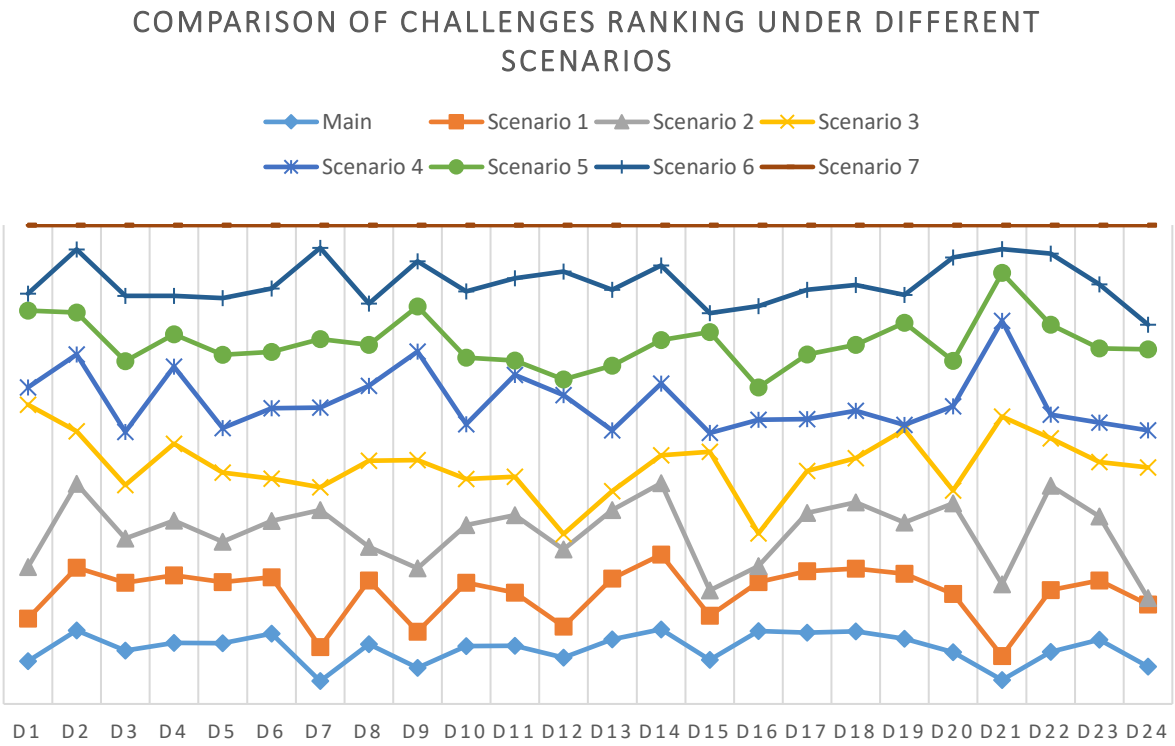


Fig 3. Ranking order based on various criterion-weighting scenarios

5.1.2. Comparative Analysis

The main goal of this section is to provide evidence of the reliability, accuracy, and effectiveness of the proposed method. To do this, a detailed comparison was made between the results from the LMAW-DNMA method and those from other established methods, such as LMAW-TODIM and LMAW-WASPAS.

Figure 4 shows the ranking results produced by DNMA, TODIM, and WASPAS. A close look at these results reveals that challenges D21, D7, and D12 consistently ranked in the top three positions across all methods. Conversely, challenges D14 and D18 were ranked lower, at twenty-fourth and twenty-third places, respectively.

After analyzing the similar and overall rankings from WASPAS and TODIM, it is clear that the challenges identified by all three methods are very similar. This is because the top six challenges, ranked from 1 to 6, are consistent across all methods. While there are slight differences in the rankings of some other challenges, these differences are generally no more than two ranks, which is acceptable given their small size. This suggests that the proposed method is valid and reliable because it produces results similar to those from other methods.

The agreement in identifying the main challenges for implementing ChatGPT in education across all three methods—DNMA, WASPAS, and TODIM—indicates that the proposed approach is strong and trustworthy. This consistency increases confidence in the accuracy of the prioritization process and supports the reliability of the findings. Using multiple methods and obtaining similar results in this study enhances its credibility. Overall, this agreement among different methods confirms that the proposed approach is valid and can be trusted for assessing the challenges of implementing ChatGPT in education.

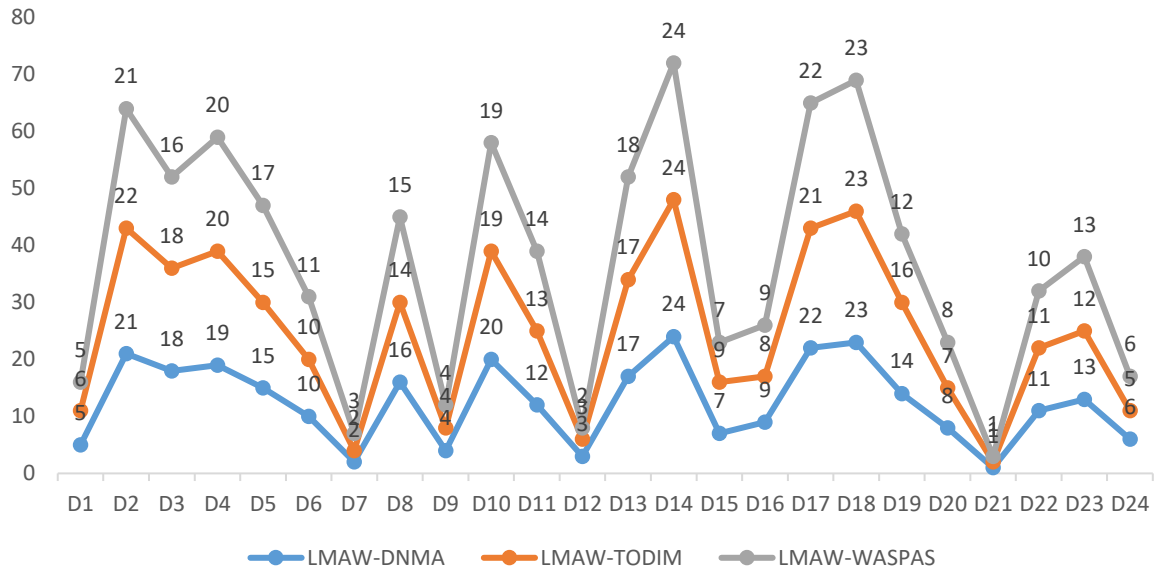


Fig 4. A comparison of the challenges' prioritization using LMAW-DNMA, LMAW-TODIM, and LMAW-WASPAS methods

5.1.3. Pearson's Correlation Analysis

The Pearson correlation coefficient is an important tool that shows how strong and in which direction two continuous variables are related [36]. In decision-making that involves multiple choices, this tool helps analyze and interpret data. It also supports DMs in choosing the best option based on certain criteria [37]. Usually, DMs must pick the best option from many choices using different criteria. The Pearson correlation helps not only in making the choice but also in comparing and checking different ranking methods used in MCDM. Different methods like DNMA, WASPAS, and TODIM can give different rankings. By calculating the Pearson correlation between these rankings, DMs can see how similar or different they are. This helps identify whether the rankings are reliable or if they need to be reviewed.

Pearson's correlation is useful for checking the relationship between how options are scored and how they are ranked. If the results from different methods are similar, it means the process is dependable [38]. If they differ a lot, it might be a sign to revisit the criteria or scoring method. By calculating the correlation between different ranking methods, DMs can see which methods give similar results and which do not. A high correlation means the methods are consistent, while a low one shows significant differences [39,40]. This helps in understanding how trustworthy the rankings are. When several methods agree on the top choices, DMs can be more confident in their decisions.

The formula for Pearson's correlation coefficient is shown below [41]:

$$r = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})*(y_i-\bar{y})}{\sqrt{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right)*\left(\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2\right)}} \quad (14)$$

In the analysis shown in Figure 5, the correlations between the three ranking methods—DNMA, TODIM, and WASPAS—provide useful insights into how accurate and suitable these methods are for evaluating the challenges of implementation ChatGPT in education.

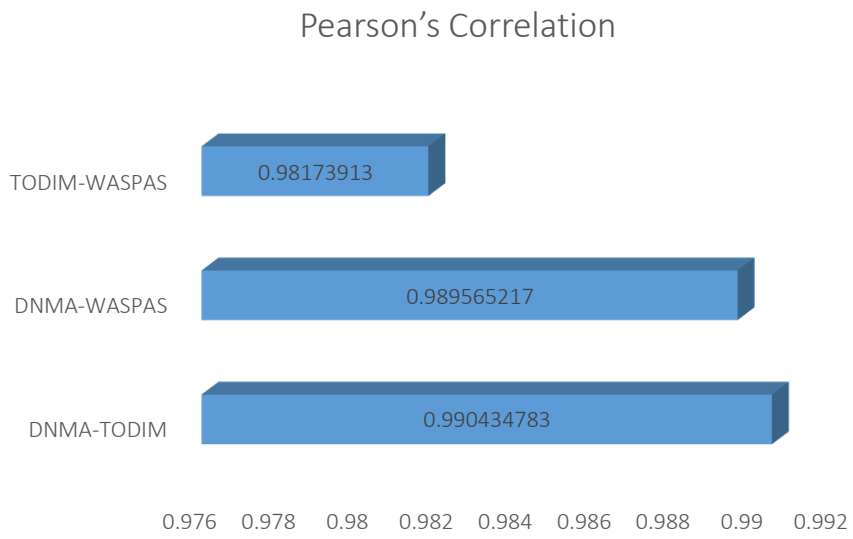


Fig 5. Pearson's correlations between DNMA, TODIM, and WASPAS

The Pearson correlation values between DNMA and the other two methods, TODIM and WASPAS, show that they produce very similar rankings. Specifically:

- A correlation of 0.9904 between DNMA and TODIM indicates that their rankings are almost identical. This means both methods evaluate the options in a very similar way when looking at the risks.
- A correlation of 0.9895 between DNMA and WASPAS also shows a strong agreement, meaning these two methods give similar ranking results.

The strong connection between DNMA and the other methods suggests that DNMA provides stable and trustworthy rankings for assessing the challenges of using ChatGPT in education. A high correlation like this shows that the method is both accurate and consistent, even when the input data changes. This reliability is important because DMs need believable and valid results.

Additionally, TODIM and WASPAS have a high correlation of 0.9817 between them, which means they work on similar principles for evaluating challenges. Since DNMA has the highest correlation with TODIM (0.9904), it appears to be a very accurate and effective method for this particular task.

From this analysis of the correlation numbers, we can say that DNMA gives reliable ranking results for assessing the challenges of implementing ChatGPT in education. Its high correlations with TODIM and WASPAS show that it agrees well with other methods and correctly captures the relationships among the challenges. Because of its consistency and strong performance, DNMA is recommended as the best method for this study. It provides accurate results that support better decision-making about the challenges of using ChatGPT in education.

The utilization of the LMAW-DNMA and Hybrid FMEA methodologies, as employed in this study, offers a well-organized and methodical approach towards the identification and prioritization of the possible drawbacks associated with the utilization of ChatGPT in an educational context. The LMAW-DNMA method, in particular, amalgamates the preferences of DMs into the analysis process, while the Hybrid FMEA approach, on the other hand, contributes towards a more comprehensive evaluation of the risks that are linked to each specific disadvantage. The outcomes derived from these findings strongly indicate that both the aforementioned methodologies showcase efficancy when it comes to the identification and prioritization of the disadvantages that are correlated with the application of ChatGPT in an educational setting.

6. Discussion and Practical Implications

The findings of this study reveal notable insights into the most critical challenges associated with the implementation of ChatGPT in educational settings. Through the application of a MCDM approach integrating hybrid FMEA, we identified both converging and diverging rankings that reflect the multifaceted risks inherent to generative AI usage in academic environments. Notably, D1 (over-reliance and laziness) ranked highest in the FMEA analysis and fifth in LMAW-DNMA, underlining the substantial behavioral concern that prolonged or unmoderated usage of ChatGPT may diminish student motivation, self-initiative, and independent learning capacity. The elevated position of D7 (negative impact on critical thinking)—ranked second in both methods—reinforces this concern and suggests that educational institutions must carefully balance AI integration with pedagogical strategies that foster cognitive engagement.

Moreover, D12 (inadequate referencing and citation) emerged as a critical challenge, securing the third position in both methodologies. This consistency indicates a systemic issue regarding the lack of verifiable source attribution in ChatGPT outputs, which poses a threat to academic integrity and may lead to misinformation. Educational stakeholders should consider embedding AI literacy programs into curricula to educate students on responsible AI usage, particularly in terms of citation practices and the evaluation of generated content. Similarly, D21 (lack of creativity and originality) achieved top ranks in LMAW-DNMA (1st) and was positioned fourth in the FMEA analysis, drawing attention to ChatGPT's potential limitations in supporting creative thinking or innovation-driven learning. This limitation may reduce the effectiveness of AI as a tool in disciplines that emphasize originality, such as literature, philosophy, and design.

Interestingly, while ethical and legal concerns such as D9 (bias, plagiarism, and data privacy issues) and D17 (copyright issues) are often emphasized in literature, their rankings were somewhat lower and more variable across methods (e.g., D9: 8th in FMEA, 4th in LMAW-DNMA; D17: 16th in FMEA, 22nd in LMAW-DNMA). This disparity may reflect a perception gap between technical criticality and the subjective severity perceived by decision-makers. Although not at the top of the list, these issues demand regulatory oversight and institutional policies to safeguard data usage, manage intellectual property concerns, and ensure fairness. Establishing clear guidelines for ChatGPT deployment, content verification mechanisms, and audit trails may help mitigate these risks.

From a practical standpoint, the divergence between the rankings across methods—such as D3 (inaccuracy of answers) being ranked 24th in FMEA and 18th in LMAW-DNMA—highlights the influence of methodological perspectives. FMEA prioritizes risk severity and potential failure impact, whereas LMAW-DNMA incorporates a broader decision-making dimension including desirability and proximity to an ideal solution. Such methodological triangulation is crucial for capturing the complexity of ChatGPT's educational implications. For instance, D15 (excessive detail in answers) ranks significantly higher in LMAW-DNMA (7th) compared to FMEA (21st), pointing to a possible disconnect between technical risk and practical usability—suggesting that the richness of information can at times hinder rather than help student comprehension.

Institutions should prioritize challenges that score consistently high across both models—such as D1, D7, D12, and D21—by fostering digital discipline, reinforcing critical thinking, and encouraging proper citation practices. Simultaneously, they should not ignore lower-ranked challenges, as the rapidly evolving nature of generative AI may amplify their relevance over time. Ultimately, the practical success of ChatGPT in education hinges not only on its technical sophistication but also on the proactive and informed governance of its implementation.

It can be concluded that the findings from this study hold great significance and relevance to educators and policymakers who are contemplating the implementation of ChatGPT within

educational environments. The utilization of natural language processing (NLP) models, like ChatGPT, has gained significant popularity within the field of education, mainly due to its ability to enrich learning outcomes and actively engage students. However, as is the case with any novel technology, it is of utmost importance to meticulously assess and manage the potential disadvantages and risks that accompany its implementation. On the whole, the outcomes derived from this study serve as a valuable source of insights into the potential risks and disadvantages that arise from the utilization of ChatGPT within an educational context. These findings are likely to prove highly beneficial for academic discourse, as well as practical decision-making, assisting educators, educational administrators, policymakers, curriculum developers, educational content creators, researchers, academic community, technology developers, AI providers, as well as other relevant stakeholders who are actively considering the integration of NLP models within educational settings.

7. Concluding Remarks and Outlook

7.1. Conclusions

This study has shed light on the drawbacks associated with the use of ChatGPT through the adoption of two distinct approaches, LMAW-DNMA and FMEA. The findings indicate that ChatGPT's lack of creativity and originality in responses, potential negative impact on the development of critical thinking skills, and inaccurate or inadequate referencing and citation are significant concerns. However, it is important to recognize that these limitations may vary depending on the context of use and the evolving capabilities of the model. The FMEA approach also highlights the risk of individuals becoming overly reliant on the technology or becoming lazier as a major issue. Nevertheless, the extent to which over-reliance occurs could be influenced by user training and implementation strategies. These findings suggest that while ChatGPT offers many potential benefits, caution and critical thinking are necessary when utilizing this technology. To address these concerns, developers could enhance ChatGPT's natural language generation capabilities, improve its referencing and citation capabilities, and incorporate features that encourage independent thinking and discourage over-reliance on the technology. Nonetheless, such improvements may face technical and ethical challenges, and their effectiveness remains to be fully validated.

In terms of methodological contribution, this study advances the field of decision analysis in AI risk assessment by systematically evaluating and comparing the effectiveness of our proposed MCDM approaches. Specifically, for validation, we compared our techniques with two well-established MCDM methods, TODIM and WASPAS. These comparisons serve to enhance the credibility and robustness of our proposed models, providing evidence of their reliability in prioritizing risks and issues related to AI technology deployment. Moreover, this research contributes to the existing body of knowledge by demonstrating how diverse MCDM techniques can be adapted and integrated into AI risk management frameworks, which is valuable for both researchers and practitioners seeking systematic decision-support tools.

7.2. Limitations of the Study

A notable limitation of this study is the limited access to real-world data, which constrained the empirical validation and reduced the generalizability of the findings. This data scarcity may have impacted the accuracy of risk assessments and the broader applicability of results across different contexts. Despite this, the comparative approach with TODIM and WASPAS offers valuable insights into the relative performance of our methods, although further validation with comprehensive datasets is necessary to solidify these contributions.

Although this study does not explicitly incorporate established technology adoption models such as the Technology Acceptance Model (TAM) or Unified Theory of Acceptance and Use of Technology (UTAUT), the underlying principles of these frameworks are inherently compatible with the application of the MCDM techniques employed herein. Both paradigms emphasize the importance of systematic evaluation of multifaceted factors—such as perceived risks, benefits, and stakeholder concerns—to facilitate informed decision-making processes. MCDM methods, including FMEA, LMAW-DNMA, TODIM, and WASPAS, operationalize this approach by ranking issues based on criteria such as detection, likelihood of occurrence, and severity, which conceptually align with constructs like perceived usefulness and perceived barriers prominent in technology acceptance theories. However, the applicability of these models directly to AI adoption in educational contexts may be limited by the variability in stakeholder perceptions, the rapid evolution of AI technology, and the constraints imposed by limited data availability. Consequently, our methodological approach offers a rigorous decision-analytic tool that can effectively complement theoretical frameworks, enabling policymakers and educational stakeholders to identify, prioritize, and address critical challenges associated with AI integration.

7.3 Directions for Future Research

For future research, we recommend incorporating these behavioral and acceptance models to further enhance understanding of the factors influencing AI adoption and to develop more comprehensive decision-making strategies in educational settings, acknowledging that data limitations may continue to pose challenges.

In summary, the use of ChatGPT presents several significant challenges that must be addressed to maximize its potential benefits while minimizing its drawbacks. While this study outlines key issues and validates the effectiveness of our proposed methods through comparison with TODIM and WASPAS, the limited availability of real-world data remains a noteworthy constraint that may influence the generalizability and accuracy of the findings. Additionally, the rapidly evolving nature of AI technology means that some limitations identified may become less relevant or change over time. Further empirical validation with broader datasets is necessary to confirm the robustness of the proposed approaches. By addressing these concerns, researchers and developers can continue to advance the capabilities of this technology and expand its potential applications across various fields, although the path forward may involve overcoming significant data-related and technical challenges.

Appendix 1

Table A1
Linear Normalization Matrix

Symbol	S	O	D	R
D1	0.8667	0.9444	0.4000	1.0000
D2	0.1333	0.1667	1.0000	0.4444
D3	0.2000	0.0556	0.0500	0.0000
D4	0.0000	0.6111	0.2000	0.1111
D5	0.5333	0.3333	0.1500	0.3333
D6	0.6667	0.4444	0.5000	0.4444
D7	0.8000	0.5000	0.8500	0.6667
D8	0.5333	1.0000	0.3500	0.2778
D9	0.4667	0.6667	0.5000	0.5000
D10	0.2667	0.0556	0.1000	0.0000
D11	0.6000	0.7778	1.0000	0.1667

Table A1
Continued

Symbol	S	O	D	R
D12	0.8000	0.6667	0.8000	0.5556
D13	0.9333	0.2222	0.7000	0.1667
D14	0.4667	0.0000	1.0000	0.0000
D15	0.4000	0.4444	0.0000	0.4444
D16	0.6000	0.5556	0.5500	0.2222
D17	0.9333	0.2222	0.4500	0.1667
D18	0.6667	0.1111	0.3000	0.1667
D19	1.0000	0.5556	0.0500	0.9444
D20	0.8000	0.4444	0.8500	0.3889
D21	0.6000	0.7222	0.6000	0.7222
D22	0.6000	0.0556	0.4000	0.5000
D23	0.8667	0.4444	0.5000	0.6111
D24	0.7333	0.8333	0.1000	0.7222

Table A2
Vector Normalization Matrix

Symbol	S	O	D	R
D1	0.9806	0.9885	0.8526	1.0000
D2	0.8742	0.8275	1.0000	0.8539
D3	0.8839	0.8045	0.7666	0.7370
D4	0.8548	0.9195	0.8035	0.7662
D5	0.9323	0.8620	0.7912	0.8247
D6	0.9516	0.8850	0.8772	0.8539
D7	0.9710	0.8965	0.9632	0.9123
D8	0.9323	1.0000	0.8403	0.8101
D9	0.9226	0.9310	0.8772	0.8685
D10	0.8935	0.8045	0.7789	0.7370
D11	0.9419	0.9540	1.0000	0.7809
D12	0.9710	0.9310	0.9509	0.8831
D13	0.9903	0.8390	0.9263	0.7809
D14	0.9226	0.7930	1.0000	0.7370
D15	0.9129	0.8850	0.7544	0.8539
D16	0.9419	0.9080	0.8895	0.7955
D17	0.9903	0.8390	0.8649	0.7809
D18	0.9516	0.8160	0.8280	0.7809
D19	1.0000	0.9080	0.7666	0.9854
D20	0.9710	0.8850	0.9632	0.8393
D21	0.9419	0.9425	0.9017	0.9270
D22	0.9419	0.8045	0.8526	0.8685
D23	0.9806	0.8850	0.8772	0.8977
D24	0.9613	0.9655	0.7789	0.9270

Table A3
Standard Deviations of Criteria

σ_1	σ_2	σ_3	σ_4
0.149159	0.199948	0.252719	0.217994

Table A4
Normalized Standard Deviation Values

w_1^{σ}	w_2^{σ}	w_3^{σ}	w_4^{σ}
0.181941	0.243893	0.308261	0.265905

Table A5
Adjusted Weight Values

$\widetilde{w}_1$	$\widetilde{w}_2$	$\widetilde{w}_3$	$\widetilde{w}_4$
0.212207	0.255025	0.266484	0.266285

Table A6
CCM, UCM and ICM Values

Symbol	CCM		UCM		ICM	
	$u_1(a_i)$	Rank	$u_2(a_i)$	Rank	$u_3(a_i)$	Rank
D1	0.797647	5	0.15989	8	0.951622	5
D2	0.455631	18	0.212521	13	0.887927	20
D3	0.349668	24	0.266285	23	0.895585	17
D4	0.390653	21	0.217869	14	0.904853	16
D5	0.612974	15	0.191535	10	0.908157	14
D6	0.759609	7	0.088762	4	0.933278	8
D7	0.825073	3	0.10501	5	0.961674	3
D8	0.535439	16	0.192317	11	0.889291	19
D9	0.803146	4	0.066621	2	0.96434	2
D10	0.365268	23	0.266285	23	0.891703	18
D11	0.636541	12	0.221904	17	0.913409	13
D12	0.876131	2	0.081365	3	0.959311	4
D13	0.520341	17	0.218734	16	0.883941	21
D14	0.365514	22	0.266285	23	0.854289	24
D15	0.712296	9	0.266484	24	0.926322	10
D16	0.791242	6	0.167661	9	0.932738	9
D17	0.448961	19	0.218734	16	0.867929	23
D18	0.4412	20	0.212521	12	0.879043	22
D19	0.618703	14	0.25316	20	0.905434	15
D20	0.721384	8	0.144455	7	0.937448	6
D21	0.918991	1	0.045097	1	0.983784	1
D22	0.63538	13	0.231411	18	0.915426	12
D23	0.684494	11	0.124243	6	0.923671	11
D24	0.704525	10	0.234506	19	0.933319	7

Acknowledgement
This research was not funded by any grant.

Conflicts of Interest
The authors declare no conflicts of interest.

References
[1] Sahoo, S. K., Choudhury, B. B., Dhal, P. R., Sahu, S., Majhi, S., & Dhar, I. (2026). Large Language Models in Multi Criteria Decision Making: A Systematic Review, Taxonomy, and Future Research Agenda. *Applied Research Advances*, 2(1), 116-136. https://doi.org/10.65069/ara21202614

[2] Kirmani, A. R. (2023). Artificial Intelligence-Enabled Science Poetry. *ACS Energy Letters*, 8(1), 574–576. https://doi.org/10.1021/acsenergylett.2c02758

[3] Tajik, E., & Tajik, F. (2023). A Comprehensive Examination of the Potential Application of Chat GPT in Higher Education Institutions. *TechRxiv*. https://doi.org/10.36227/techrxiv.22589497.v1
[4] Lee, H. (2023). The Rise of ChatGPT: Exploring Its Potential in Medical Education. *Anatomical Sciences Education*. https://doi.org/10.1002/ase.2270
[5] Verma, M. (2023). Novel Study on AI-Based Chatbot (ChatGPT) Impacts on the Traditional Library Management. *International Journal of Trend in Scientific Research and Development (IJTSRD)*, 7(1), 961–964. www.ijtsrd.com/papers/ijtsrd52767.pdf
[6] Aithal, S., & Aithal, P. S. (2023). Effects of AI-Based ChatGPT on Higher Education Libraries. *International Journal of Management, Technology, and Social Sciences*, 95–108. https://doi.org/10.47992/IJMTS.2581.6012.0272
[7] Lund, B. D., & Wang, T. (2023). Chatting about ChatGPT: How May AI and GPT Impact Academia and Libraries? *Library Hi Tech News*, 40(3), 26–29. https://doi.org/10.1108/LHTN-01-2023-0009
[8] Tai, M. C.-T. (2020). The Impact of Artificial Intelligence on Human Society and Bioethics. *Tzu Chi Medical Journal*, 32(4), 339. https://doi.org/10.4103/tcmj.tcmj_71_20
[9] Wogu, I. A. P., Olu-Owolabi, F. E., Assibong, P. A., Agoha, B. C., Sholarin, M., Elegbeleye, A., Igbokwe, D., & Apeh, H. A. (2017). Artificial Intelligence, Alienation and Ontological Problems of Other Minds: A Critical Investigation into the Future of Man and Machines. In *2017 International Conference on Computing Networking and Informatics (ICCNI)* (pp. 1–10). IEEE. https://doi.org/10.1109/ICCNI.2017.8123792
[10] Yilmaz, R., & Karaoglan Yilmaz, F. G. (2023). Augmented Intelligence in Programming Learning: Examining Student Views on the Use of ChatGPT for Programming Learning. *Computers in Human Behavior: Artificial Humans*, 1(2), 100005. https://doi.org/10.1016/j.chbah.2023.100005
[11] Sallam, M. (2023). ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. *Healthcare*, 11(6), 887. https://doi.org/10.3390/healthcare11060887
[12] Gill, S. S., Xu, M., Patros, P., Wu, H., Kaur, R., Kaur, K., ... & Buyya, R. (2024). Transformative effects of ChatGPT on modern education: Emerging Era of AI Chatbots. *Internet of Things and Cyber-Physical Systems*, *4*, 19-23.. https://doi.org/10.1016/j.iotcps.2023.06.002
[13] Adiguzel, T., Kaya, M. H., & Cansu, F. K. (2023). Revolutionizing Education with AI: Exploring the Transformative Potential of ChatGPT. *Contemporary Educational Technology*, 15(3), ep429. https://doi.org/10.30935/cedtech/13152
[14] Zhang, B. (2023). Preparing Educators and Students for ChatGPT and AI Technology in Higher Education: Benefits, Limitations, Strategies, and Implications of ChatGPT & AI Technologies. Unpublished. https://doi.org/10.13140/RG.2.2.32105.98404
[15] Malinka, K., Peresíni, M., Firc, A., Hujnák, O., & Janus, F. (2023). On the Educational Impact of ChatGPT: Is Artificial Intelligence Ready to Obtain a University Degree? In *Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1* (pp. 47–53). ACM. https://doi.org/10.1145/3587102.3588827
[16] Chukwuere, J. E. (2024). The use of ChatGPT in higher education: The advantages and disadvantages (arXiv:2403.19245) [Preprint]. *arXiv*. https://doi.org/10.48550/arXiv.2403.19245
[17] Li, L., Ma, Z., Fan, L., Lee, S., Yu, H., & Hemphill, L. (2023). ChatGPT in education: A discourse analysis of worries and concerns on social media. *Education and Information Technologies*. https://doi.org/10.1007/s10639-023-12256-9
[18] Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2024). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International*, 61(3), 460–474. https://doi.org/10.1080/14703297.2023.2195846
[19] Montenegro-Rueda, M., Fernández-Cerero, J., Fernández-Batanero, J. M., & López-Meneses, E. (2023). Impact of the Implementation of ChatGPT in Education: A Systematic Review. *Computers*, 12(8), 153. https://doi.org/10.3390/computers12080153
[20] Nayeb-Pashaei, K., Vahabzadeh, S., Hosseinian-Far, A., & Ghoushchi, S. J. (2026). Sustainable Urban Transportation: Key Criteria for a Greener Future. *Spectrum of Operational Research*, 103-127. https://doi.org/10.31181/sor31202634
[21] Sarkar, A., Goswami, S. S., Behera, D. K., & Bozanic, D. (2026). Criteria Weighting Methods in Multi-Criteria Decision Making: A Comprehensive Review of Subjective, Objective, and Hybrid Approaches. *Journal of Contemporary Decision Science*, 3(1), 1-34. https://doi.org/10.67334/cds202619
[22] Sarkar, A., & Goswami, S. S. (2026). A Review of the Application of MCDM Methods in Business Analytics. *Applied Decision Analytics*, 2(1), 150-180. https://doi.org/10.66972/ada21202614
[23] Bowles, J. B., & Peláez, C. E. (1995). Fuzzy Logic Prioritization of Failures in a System Failure Mode, Effects and Criticality Analysis. *Reliability Engineering and System Safety*, 50(2), 203–213. https://doi.org/10.1016/0951-8320(95)00068-D

[24] Mohammadi, M., Sarvi, S., & Jafarzadeh Ghoushchi, S. (2025). Assessing and Prioritizing Construction Contracting Risks with an Extended FMEA Decision-Making Model in Uncertain Environments. *Spectrum of Decision Making and Applications*, 3(1), 187-211. https://doi.org/10.31181/sdmap31202642

[25] Vahabzadeh, S., Haghshenas, S. S., Ghoushchi, S. J., Guido, G., Simic, V., & Marinkovic, D. (2025). A new framework for risk assessment of road transportation of hazardous substances. *Facta Universitatis, Series: Mechanical Engineering*. https://doi.org/10.22190/FUME240801007V

[26] Maghami, M. R., Vahabzadeh, S., Mutambara, A. G. O., Ghoushchi, S. J., & Gomes, C. (2024). Failure analysis in smart grid solar integration using an extended decision-making-based FMEA model under uncertain environment. *Stochastic Environmental Research and Risk Assessment*, 38(9), 3543-3563. https://doi.org/10.1007/s00477-024-02764-6

[27] Ghoushchi, S. J., Haghshenas, S. S., Vahabzadeh, S., Guido, G., & Geem, Z. W. (2024). An integrated MCDM approach for enhancing efficiency in connected autonomous vehicles through augmented intelligence and IoT integration. *Results in Engineering*, 23, 102626. https://doi.org/10.1016/j.rineng.2024.102626

[28] Ghoushchi, S. J., Vahabzadeh, S., & Pamucar, D. (2024). Applying hesitant q-rung orthopair fuzzy sets to evaluate uncertainty in subsidence causes factors. *Heliyon*, 10(8). https://doi.org/10.1016/j.heliyon.2024.e29415

[29] Khalid, I., & Riaz, L. (2025). Behavioral Biases and Investment Management Decisions: The Mediating Role of Risk Perception and the Moderating Role of Financial Literacy. *Management Science Advances*, 3(1), 1-19. https://doi.org/10.31181/msa31202629

[30] Pamučar, D., Žižović, M., Biswas, S., & Božanić, D. (2021). A NEW LOGARITHM METHODOLOGY OF ADDITIVE WEIGHTS (LMAW) FOR MULTI-CRITERIA DECISION-MAKING: APPLICATION IN LOGISTICS. *Facta Universitatis, Series: Mechanical Engineering*, 19(3), 361. https://doi.org/10.22190/FUME210214031P

[31] Tesić, D., Božanić, D., Pamučar, D., Miljković, B., & Puška, A. (2025). MCDM model for the selection of network planning techniques in the army for the purposes of performing engineering works when overcoming water obstacles. *Journal of Decision Analytics and Intelligent Computing*, 5(1), 70-86. https://doi.org/10.31181/jdaic10006062025t

[32] Liao, H., & Wu, X. (2020). DNMA: A Double Normalization-Based Multiple Aggregation Method for Multi-Expert Multi-Criteria Decision Making. *Omega*, 94, 102058. https://doi.org/10.1016/j.omega.2019.04.001

[33] Jafarzadeh Ghoushchi, S., Shaffiee Haghshenas, S., Vahabzadeh, S., Shaffiee Haghshenas, S., Astarita, V., & Guido, G. (2025). Risk assessment of young driver behavior using an extended decision-making approach based on FMEA in uncertain environments. *Neural Computing and Applications*, 1-38. https://doi.org/10.1007/s00521-025-11087-8

[34] Gomes, L. F. A. M., & Lima, M. M. P. P. (1991). TODIM: basics and application to multi-criteria ranking of projects with environmental impacts. *Foundations of Computing and Decision Sciences*, 16, 113.

[35] Chakraborty, S., & Zavadskas, E. K. (2014). Applications of WASPAS method in manufacturing decision making. *Informatica*, 25(1), 1-20. https://doi.org/10.15388/Informatica.2014.01

[36] Tufan, D., & Ulutaş, A. (2025). Supplier Selection in the Food Sector: An Integrated Approach Using LODECI and CORASO Methods. *Spectrum of Decision Making and Applications*, 3(1), 40-51. https://doi.org/10.31181/sdmap31202631

[37] Bhaduri, P., Biswas, A., Gazi, K. H., & Mondal, S. P. (2026). An MCDM-based Optimization Approach for Identifying Efficient Renewable Energy Sources under Fermatean Fuzzy Environment. *Applied Decision Analytics*, 2(1), 320-351. https://doi.org/10.66972/ada21202627

[38] Bouraima, M. B. (2026). Unveiling the Challenges of Artificial Intelligence Use in Auditing: A Holistic Multi-Criteria Decision-Making Approach. *Applied Research Advances*, 2(1), 137-145. https://doi.org/10.65069/ara21202613

[39] Biswas, S., Biswas, B., Sanyal, A., Chatterjee, P., & Pamucar, D. (2026). A Multi-Stage Fuzzy Decision Analytics Framework for Evaluating Climate-Induced Risk Factors of Catastrophic Healthcare Expenditure. Journal of Contemporary Decision Science, 3(1), 1-35. https://doi.org/10.67334/cds31202628

[40] Sarkar, A., Goswami, S. S., Behera, D. K., & Bozanic, D. (2026). Criteria Weighting Methods in Multi-Criteria Decision Making: A Comprehensive Review of Subjective, Objective, and Hybrid Approaches. Journal of Contemporary Decision Science, 3(1), 1-34. https://doi.org/10.67334/cds202619

[41] Basuri, T., Gazi, K. H., Bhaduri, P., Das, S. G., & Mondal, S. P. (2025). Decision-analytics-based Sustainable Location Problem - Neutrosophic CRITIC-COPRAS Assessment Model. *Management Science Advances*, *2*(1), 19-58. https://doi.org/10.31181/msa2120257