ISSN: 2583-9888
DOI: doi.org/10.51219/JAIMLD/mohammed-sallam/20

URF PUBLISHERS
connect with research world

Journal of Artificial Intelligence, Machine Learning and Data Science

https://urfpublishers.com/journal/artificial-intelligence

Vol: 1 & Iss: 4

Research Article

Assessment of Artificial Intelligence Credibility in Evidence-Based Healthcare Management with "AERUS" Innovative Tool

Dr. Mohammed Sallam[1,2*], **Dr. Johan Snygg**[3,4] **and Dr. Malik Sallam**[5,6,7,]

[1]School of Business, International American University, Los Angeles, California, USA

[2]Department of Pharmacy, Mediclinic Parkview Hospital, Mediclinic Middle East, Dubai, UAE

[3]Mediclinic City Hospital, Mediclinic Middle East, Dubai, UAE

[4]Sahlgrenska Academy at the University of Gothenburg and Sahlgrenska University Hospital, Gothenburg, Sweden

[5]Department of Pathology, Microbiology, and Forensic Medicine, School of Medicine, University of Jordan, Amman, Jordan

[6]Department of Clinical Laboratories and Forensic Medicine, Jordan University Hospital, Amman, Jordan

[7]Department of Translational Medicine, Faculty of Medicine, Lund University Malmo, Sweden

Citation: Sallam M, Snygg J, Sallam M. Assessment of Artificial Intelligence Credibility in Evidence-Based Healthcare Management with "AERUS" Innovative Tool. *J Artif Intell Mach Learn & Data Sci 2024*, 1(4), 9-18. DOI: doi.org/10.51219/JAIMLD/mohammed-sallam/20

Received: 07 February, 2024; **Accepted:** 16 February, 2024; **Published:** 19 February, 2024

***Corresponding author:** Dr. Mohammed Sallam, Department of Pharmacy, Mediclinic Parkview Hospital, Mediclinic Middle East, Dubai, UAE, mohammed.sallam@mediclinic.ae

ABSTRACT

Background: Artificial Intelligence (AI) technologies have found applications across various arenas, and Evidence-Based Management (EBMgt) is no exception. However, in this context, the careful assessment of AI outcomes becomes essential to confirm their credibility and ensure adherence to ethical standards. Mirroring recent similar explorations in the literature, this study introduced the "AERUS" tool, designed to evaluate AI trustworthiness in healthcare administration, focusing on five key areas: Accuracy, Efficiency, Reliability, Usability, and Security.

Methods: The AERUS instrument evaluated AI's reliability in healthcare administration. It underwent minor revisions after an initial test with thirty healthcare administrators and internal consistency confirmation via Cronbach's alpha. The final version was tested on four AI models (ChatGPT 3.5, ChatGPT 4, Microsoft Bing, Google Bard) over six managerial topics, with evaluation by two raters using Cohen's kappa.

Results: The refined AERUS tool assessed five areas: AI accuracy in management data, operational efficiency impact, decision-making reliability, user-friendliness for managers, and security protocol adherence. Initial testing with ten healthcare management statements showed high internal consistency (Cronbach's alpha of .911). Among six assessments, Microsoft Bing scored highest (mean 22.93, SD 1.11), followed by ChatGPT-4 (mean 22.00, SD 1.21), ChatGPT-3.5 (mean 20.00, SD 1.21), and Google Bard (mean 19.60, SD 1.22). Inter-rater agreement resulted in Cohen's kappa values ranging from 0.358 to 0.885 for the AI models.

Conclusions and Recommendations: AERUS presents a supporting instrument for addressing AI credibility concerns in EBMgt, with recommendations for further research and widespread implementation to ensure the trustworthiness and reliability of AI in professional managerial decision-making.

Keywords: Artificial Intelligence, AI, Evidence-Based Management, EBMgt, Healthcare Administrators, Decision-making, AERUS Tool, AI Credibility

1 Introduction

Considering its extensive potential and influence in decision-making, advanced artificial intelligence (AI) technologies have been widely used in various domains, including education, business, healthcare, and others. The rapid adoption of AI in the last few years highlighted its significant advantages and capabilities. While AI systems offer substantial benefits, they also carry the risk of unintended consequences for individuals and society at large[1]. Shrestha, Ben-Menahem and von Krogh[2] expressed that managers who utilize AI in making decisions remain accountable for the outcomes of those decisions. Therefore, it is crucial to thoroughly understand AI systems' limitations when applying them in business contexts, including making organizational decisions[3].

Decision-making involves gathering and assessing information from multiple sources in a given field. It is particularly complex for managers in Healthcare, involving cooperation among healthcare professionals, investors, governments, policymakers, and patients[4].

According to Barends and Rousseau[5], Evidence-Based Management (EBMgt) is best defined as making decisions through a diligent, explicit, and judicious approach using the best available evidence from various sources. This process starts with asking, which involves translating a practical issue or problem into an answerable question. Next is acquiring, which entails systematically searching for and retrieving relevant evidence. The third step is appraising, where the trustworthiness and relevance of the evidence are critically judged. Aggregating follows, involving the weighing and combining of the evidence. The fifth step is applying, where the evidence is incorporated into the decision-making process. Finally, assessing is conducted, evaluating the outcome of the decision to enhance the likelihood of achieving a favorable result.

EBMgt systematically uses the best available evidence to inform decisions[6-8]. The literature on EBMgt in healthcare organizations has consistently demonstrated the importance of its principles for improving quality and safety in Healthcare[9,10]. To ensure effective decision-making, healthcare managers require tools to categorize sources of evidence and enable critical evaluation based on shared characteristics[11]. In addition, leveraging intelligent technologies in strategic and critical decisions could significantly enhance management approaches and the quality of healthcare services provided[12,13].

AI's successful development and implementation require high-quality governance to protect data and frameworks that support the creation of trustworthy AI products[14,15]. To date, the speedy advancement of Artificial Intelligence in handling complex tasks has led to many AI-powered devices and services gaining increased autonomy in decision-making, affecting both individuals and organizations[16]. This is particularly evident for healthcare managers and providers, where AI's potential to revolutionize the industry is met with the critical challenge of ensuring the credibility and reliability of AI-generated informatio[17].

As an innovation, AI's efficiency will be primarily determined by its adoption and utilization[18]. Healthcare organizations increasingly depend on AI for critical decision-making; thus, a systematic approach and tools to assess and enhance AI reliability and credibility becomes crucial[1,19-22].

Until now, there has not been a specific tool designed to test the quality of AI-produced content for use by healthcare managers in their decision-making processes. This lack of a specific evaluation tool means that the reliability of such AI-produced content is not fully assured[23]. The AERUS tool was developed to address this gap. It can check the precision and reliability of AI-generated data, thereby assisting professionals and leaders in making more informed decisions.

1.1. Purpose and value

This research was designed to pioneer the systematic evaluation of AI credibility within EBMgt[24,25]. AI credibility means its ability to interact with individuals obviously, relevantly, consistently, empathetically, responsively, and truthfully[26]. This broad description of AI credibility is especially critical in areas where AI interfaces directly with users, such as customer service, healthcare management, and education. Additionally, the AERUS tool was introduced, and its application was piloted to enhance AI's role in healthcare managers' decision-making and systematically assess the trustworthiness of AI-generated data[21]. (Figure 1) illustrates the foundational brainstorming process that fostered the development of the AERUS tool framework.

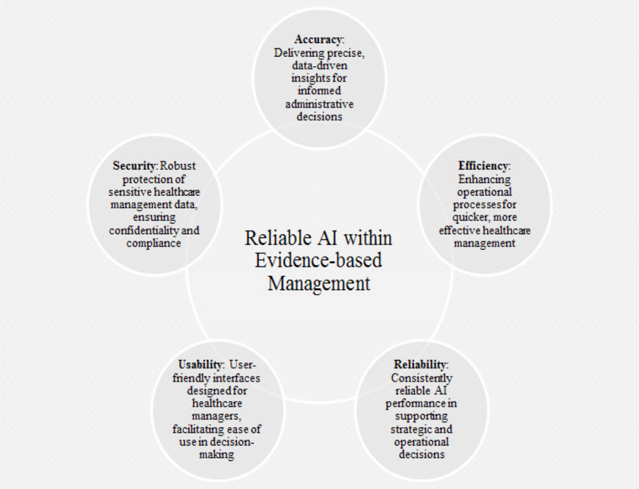
Accuracy: Delivering precise, data-driven insights for informed administrative decisions
Efficiency: Enhancing operational processes for quicker, more effective healthcare management
Reliability: Consistently reliable AI performance in supporting strategic and operational decisions
Usability: User-friendly interfaces designed for healthcare managers, facilitating ease of use in decision-making
Security: Robust protection of sensitive healthcare management data, ensuring confidentiality and compliance
Reliable AI within Evidence-based Management

Figure 1: Key factors of reliable AI in Evidence-Based Healthcare Management.

1.2. Implications on healthcare management

The effective development and deployment of the AERUS tool could significantly impact healthcare management. This tool's ability to verify the accuracy and trustworthiness of AI-driven data enables healthcare leaders and practitioners to make better-informed choices. This could lead to improved management practices, more effective resource utilization, and enhanced healthcare organizational outcomes[21,27].

2. Literature Review

2.1. Theoretical frameworks

The current study aligned the development and application of the AERUS tool with three theoretical frameworks (Table 1) to ensure the instrument's relevance and effectiveness in the study scope. The Technology Acceptance Model (TAM) by Fred D. Davis[28] informed our understanding of how admin professionals might accept and use the AERUS tool, focusing on its perceived usefulness and ease of use in evaluating AI-generated information. Additionally, the principles of AI Ethics were central to our approach, ensuring the tool aids in developing and using AI systems that are fair, accountable, and transparent[29,30].

This pathway was associated with the primary goal of building trust in AI applications within Healthcare managerial decisions. Ultimately, the principles of EBMgt guided our methodology and ensured the AERUS tool facilitated decision-making grounded in empirical data and scientific evidence, which is crucial for ethical and effective healthcare management[31,32].

Table 1: Literature-based theoretical models.

S. N.	Theoretical Framework	Reference	Contribution to the Study	Relation to AI Credibility, Transparency, and Reliability in Healthcare Management
1	Technology Acceptance Model (TAM)	Hu, Chau, Sheng and Tam[33]	Centers on the adoption and utilization of technology by users based on the perceived benefits and simplicity of use	Assesses how healthcare administrators accept and use the AERUS tool
2	AI Ethics	Patel[29] Stahl and Eke[30]	Provides principles and guidelines for ethical AI development and use, focusing on issues like fairness, accountability, and transparency	Ensures that AI systems in Healthcare are developed and used ethically, promoting trust and accountability
3	Evidence-Based Management (EBMgt)	Rousseau and McCarthy[32] Guo, Berkshire, Fulton and Hermanson[31]	Advocates for managerial decision-making grounded in scientific evidence and rational analysis, emphasizing the use of empirical data.	Aligns with assessing the application of AI in presenting evidence-based and ethical decisions in healthcare management.

2.2. Artificial Intelligence (AI) and robotic process automation (RPA)

AI and RPA promise to significantly simplify life, from helping with basic activities such as scheduling appointments, processing patient records, and transcribing doctors' notes to more complex tasks like analyzing large sets of healthcare data for strategic planning and implementing advanced patient care models. Both innovations can play a crucial role in reducing the effects of human error on an organization's workflow[34]. While AI cannot replace a human, it does offer the benefit of examining extensive data and detecting patterns that might elude the human mind[35]. (Figure 2) showcases AI/RPA applications within the healthcare sector-ranging from diagnostics, patient engagement, and treatment protocols to administrative tasks, medical imaging, and health monitoring[36]. This diagram illustrates the many-sided impact of AI in Healthcare, highlighting its potential to revolutionize each aspect by providing enhanced accuracy, efficiency, and personalized patient care.

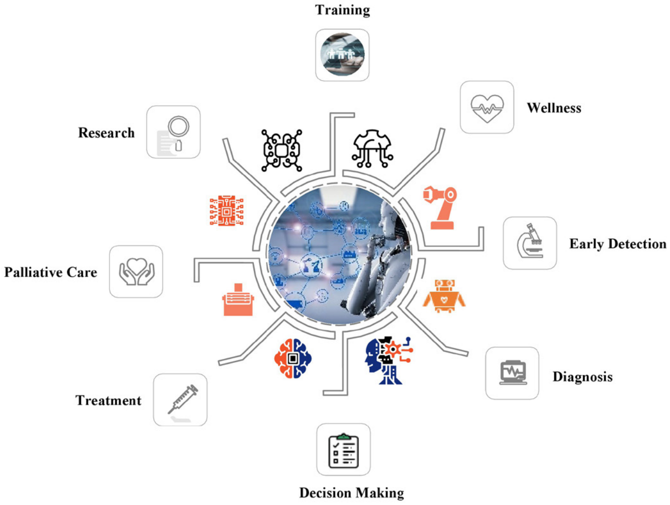

Figure 2: AI and Robotics use examples in Healthcare. Adapted from Deo and Anjankar[36].

Haleem, Javaid, Singh, Rab and Suman[37] conveyed that the outcome measures of executing AI and automation in management systems are obvious, leading to profound enhancements in operational efficiency and proven decision-making[38]. (Figure 3) illustrates the effects of industrialization (automation) on cost reduction, administrative staff productivity, process speed, patient experience, and data accuracy[39].

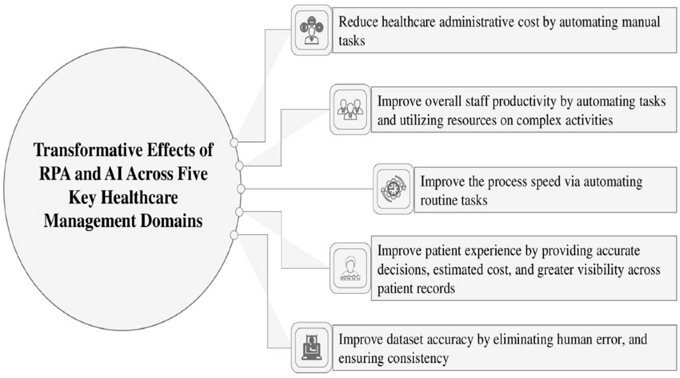
Transformative Effects of RPA and AI Across Five Key Healthcare Management Domains
Reduce healthcare administrative cost by automating manual tasks
Improve overall staff productivity by automating tasks and utilizing resources on complex activities
Improve the process speed via automating routine tasks
Improve patient experience by providing accurate decisions, estimated cost, and greater visibility across patient records
Improve dataset accuracy by eliminating human error, and ensuring consistency

Figure 3: Impact of AI and Process Automation in Healthcare Management.

2.3. Utilizing AI in the managerial decision-making process within healthcare organizations

The successful integration of artificial intelligence into strategic decision-making processes is crucial for the future competitiveness of organizations and their leaders[40]. The application and disputes of Artificial Intelligence (AI) in the managerial decision-making process within healthcare organizations have earned substantial attention in recent literature[41]. Researchers explored mixed views on the potential of AI-driven tools and algorithms to enhance decision-making efficiency and effectiveness in healthcare settings. Peer-reviewed studies highlighted AI's ability to analyze vast amounts of patient data, assisting clinicians and managers in making data-informed decisions regarding treatment plans, resource allocation, and patient outcomes[42]. Furthermore, studies emphasized the importance of ensuring AI's structural and ethical use in healthcare decision-making, addressing bias and data privacy[36,43].

2.4. AI credibility and ethical framework considerations

Recent developments in Artificial Intelligence have enabled machines to replicate various human capabilities, such as perception, emotion detection and comprehension, conversation, and even aspects of creativity[44]. However, several leading AI experts, alongside notable figures, have expressed concerns in various media outlets about the potential of a security breach associated with AI[45]. Anxiety about AI can significantly hinder its integration into workplace environments[46]. AI cannot assess the ethics of a decision on its own. It requires close oversight and extra precautions to prevent the inclusion of human biases or unfair data in its machine-learning algorithms[47]. Furthermore, AI falls short in terms of creativity, a trait inherent to humans, and the capacity to experience emotions and engage in critical thinking, which is essential in decision-making[48].

AI credibility and ethical considerations are vital in healthcare management decisions[49]. As AI becomes increasingly integral to management decision-making, ensuring the reliability and ethical integrity of AI-driven decisions is imperative. Healthcare corporate leaders should prioritize the transparency and trustworthiness of AI systems that impact critical decisions related to patient care, resource allocation, and operational efficiency[50]. For instance, an organization with insufficient privacy policies and security measures for clients could lead to a data breach. This scenario might result in a substantial decrease in customer base, a loss of trust, reduced competitiveness in attracting employees, and a decline in stock market values[51]. Ethical concerns encompass safeguarding data privacy, mitigating algorithmic bias in decision recommendations, and addressing ethical dilemmas in care and other managerial decisions[52]. Tackling these ethical considerations and identifying competencies while upholding the credibility of AI-driven healthcare management is vital to maintaining trust among all stakeholders[53]. The above highlights the pressing need to establish ethical frameworks and standards that guide AI's responsible and beneficial use in healthcare management decision-making processes[54]. (Table 2) outlines the ethical criteria for AI practice, covering Fairness (avoiding discrimination), Transparency (making AI understandable), Accountability (responsibility for AI systems), Privacy (data protection and consent), Bias Mitigation (preventing unfair outcomes), Safety and Reliability (safe, reliable system design), Robustness (handling unexpected situations), Human Control (oversight in critical tasks), Social Impact (considering AI's societal effects), and Ethical Use (promoting beneficial AI use and avoiding unethical applications).

Table 2: Description of ethical practice criteria for AI.

Ethical Criteria	Description
Fairness	Avoid discrimination based on protected characteristics.
Transparency	Make AI algorithms and decisions understandable.
Accountability	Establish clear responsibility for AI systems.
Privacy	Protect individuals' data and obtain informed consent.
Bias Mitigation	Identify and mitigate bias to prevent unfair outcomes.
Safety and Reliability	Design systems to operate safely and reliably.
Robustness	Ensure AI can handle unexpected situations and attacks.
Human Control	Maintain human oversight, especially in critical tasks.
Social Impact	Consider broader societal implications of AI deployment.
Ethical Use	Promote the beneficial use of AI and avoid unethical purposes.

Source of information tailored from Floridi, Cowls, Beltrametti, Chatila, Chazerand, Dignum, Luetge, Madelin, Pagallo, Rossi, Schafer, Valcke and Vayena[55].

3. Methods

3.1. Study design

The study was tailored from the methodologies encapsulated in the CLEAR tool[21]. This tool examined existing approaches for evaluating information quality to establish a framework suitable for appraising data generated by AI-based systems. In the context of healthcare management, the innovative AERUS instrument was employed to analyze AI's (1) Accuracy, (2) Efficiency, (3) Reliability, (4) Usability, and (5) Security. A thorough analysis informed the development of AERUS of data sources, a careful assessment of the transparency and performance of diverse AI models, and the incorporation of end-user feedback to ensure the tool's adaptability and relevance[21] (Table 3).

Table 3: AERUS tool five components description.

AERUS Step	Description	Question 1	Question 2
Accuracy (A)	Assessing the accuracy of AI in reflecting managerial data and strategies	How accurately does the AI model reflect key performance indicators and management strategies?	How does the AI system ensure data accuracy in strategic planning and forecasting?
Efficiency (E)	Evaluating how AI enhances operational efficiency in healthcare management	How does the AI system improve operational efficiency in resource allocation or scheduling areas?	Can the AI system effectively streamline administrative processes, reducing time and cost?
Reliability (R)	Assessing the consistency of AI in supporting managerial decisions	How consistently does the AI provide reliable support for various administrative and managerial decisions?	Does the AI system maintain its performance reliability during critical healthcare operations and decision-making?
Usability (U)	Determining the ease of use of AI systems for healthcare managers	How user-friendly is the AI system for managers to incorporate into their daily administrative tasks?	Is the AI interface intuitive for healthcare managers, enabling quick adaptation and efficient use?
Security (S)	Examining AI compliance with healthcare management data security standards	How effectively does the AI system protect sensitive data and adhere to security protocols?	Does the AI system have robust measures to prevent data breaches and ensure the confidentiality of healthcare information?

3.2. Methods appropriateness

3.2.1. Evaluation of the content validity of the AERUS instrument

The content validity for the AERUS tool involved seeking input from three healthcare executives. All leader's recommendations and suggestions were utilized[21].

3.2.2. Testing of the AERUS tool

The research involved a diverse selection of potential participants from different healthcare organizations. This varied group was selected to offer a complete viewpoint on the research subject, covering various leadership and managerial roles. The total number of healthcare managers who offered feedback was thirty, distributed between Executives, Directors, Senior Managers, Managers, Middle Managers, Unit Managers, Junior Managers, and Supervisors.

The initial test evaluated ten statements about Evidence-Based Healthcare Management using the AERUS tool. The statements for evaluation were created based on discussions in an expert forum, encompassing diverse management topics to confirm the tool's preliminary relevance for a wide array of subjects within EBMgt. The statements were crafted to contain correct and incorrect information, incorporating irrelevant, inaccurate, or vague content in a purposeful but randomized manner. The statements evaluated using the AERUS tool are detailed in (Table 4).

Table 4: AERUS tool ten statements.

S. No.	Statement	Accuracy	Inclusion of Irrelevant/Ambiguous Content
1	Utilizing predictive analytics can decrease patient no-show rates by 25%.	Accurate	None
2	Employee satisfaction increases by 10% when meetings are scheduled bi-weekly instead of weekly.	Inaccurate	Deliberate Irrelevance
3	Introducing an AI system can reduce prescription medication errors by up to 40%.	Accurate	None
4	Upgrading the cafeteria menu correlates with a 15% improvement in patient recovery rates.	Inaccurate	Vague Content
5	Telemedicine visits have been proven to reduce hospital readmission rates by 30%.	Accurate	None
6	A 20% budget increase for departmental marketing will lead to a 50% rise in patient admissions.	Inaccurate	Deliberate Irrelevance
7	Automated HR systems will save up to 60 hours of manual work per month.	Accurate	None
8	Switching to LED lighting in all facilities will improve the accuracy of clinical diagnoses.	Inaccurate	Vague Content
9	Frequent team-building sessions have been associated with a 5% decrease in medical errors.	Inaccurate	Deliberate Irrelevance
10	Implementing electronic health records has increased patient data access by 80%.	Accurate	None

Every participant was asked to evaluate AERUS's five elements using a 5-point Likert scale, ranging from (5) excellent to (1) poor[21].

3.2.3. Completion and practical implementation of the AERUS tool for evaluating widely used AI-based applications

Post refinement, informed by pilot feedback, we deployed the AERUS instrument to appraise content generated by six standard management-related inquiries across several AI interfaces, including ChatGPT 3.5, ChatGPT-4, Bing, and the Google Bard (Table 5). The selection of the four AI models was primarily guided by their significant applicability and integration potential within the domain of healthcare management. These models stand out for their advanced data processing, analysis, and decision support capabilities, which are critical functionalities in this sector. Their prominence in the industry and their proven effectiveness in handling complex data make them ideal candidates for evaluating the efficacy of the AERUS tool in a real-world healthcare management context. This choice ensures that the study's findings are theoretically robust and practically relevant to the evolving landscape of AI-assisted healthcare administration.

Additionally, the study's blend of the six questions was strategically intended to capture a valuable spectrum of crucial healthcare management areas, including patient care, staff management, operational efficiency, and strategic planning. These questions, rooted in evidence-based healthcare literature, accurately reflect the diverse and real-world challenges faced in the healthcare sector. They were chosen not only for their breadth and relevance but also for their ability to demonstrate the practical impact and adaptability of AI in various sides of decision-making.

A new conversation thread was initiated for each interface in every response, maintaining uniform prompts throughout the evaluation. Two senior evaluators independently assessed the content generated by the AI model[21].

3.3. Statistical analysis

Statistical analysis was performed using IBM SPSS Version 29.0 for Windows, with a significance level set at a *P*-value less than 0.05. Descriptive statistical analysis used the mean and standard deviation (SD). The internal consistency of the AERUS tool was assessed using Cronbach's alpha[56].

Table 5: AERUS tool six questions for evaluating widely used AI-based applications.

S. No.	Inquiry Question
1	Does predictive analytics reduce patient no-shows?
2	Do frequent meetings affect staff satisfaction?
3	Can AI lower medication errors?
4	Is patient recovery linked to cafeteria menus?
5	Does telemedicine cut readmission rates?
6	Does a higher marketing budget boost patient admission?

The final score for the AERUS tool was calculated by summing up the average scores given by two independent raters for each item, ranked as excellent = 5, very good = 4, good = 3, satisfactory/fair = 2, or poor = 1. The cumulative AERUS scores, which vary from 6 to 30, have been randomly divided into three tiers: (1) "Limited" content for scores from 6 to 13, (2) "Adequate" content for scores between 14 and 21, and (3) "Superior" content for scores from 22 to 30 (Table 6).

Table 6: The score for the AERUS tool.

AERUS Score Range	(1-5) Likert Rating per Item	Category
6-13	1 (Poor) to 2 (Fair)	Limited
14-21	3 (Good) to 4 (Very Good)	Adequate
22-30	5 (Excellent)	Superior

3.4. Ethical considerations

3.4.1. Confidentiality

The research took great care to guarantee the confidentiality of the information collected by the pilot, with a dedication to safeguarding the privacy of every respondent and evaluator.

3.4.2. Institutional review board (IRB)

Given the evaluative nature of the study design, IRB was not required.

4. Results

4.1. The completed items of the AERUS tool

Figure 4 displays the finalized wording of the AERUS tool items. Each dimension is critical to the overall credibility of AI applications in management settings, indicating the tool's comprehensive approach to evaluating AI systems.

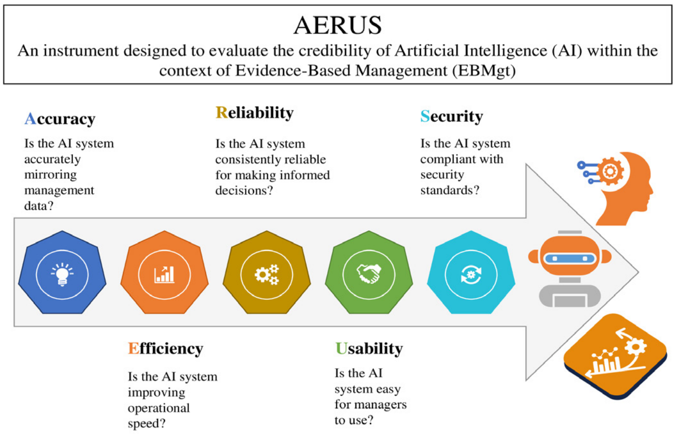
AERUS
An instrument designed to evaluate the credibility of Artificial Intelligence (AI) within the context of Evidence-Based Management (EBMgt)
Accuracy
Is the AI system accurately mirroring management data?
Reliability
Is the AI system consistently reliable for making informed decisions?
Security
Is the AI system compliant with security standards?
Efficiency
Is the AI system improving operational speed?
Usability
Is the AI system easy for managers to use?

Figure 4: The finalized wording of the AERUS tool items.

4.2. Outcomes from the initial trial run of the AERUS instrument

(Table 7) the AERUS tool's initial testing phase, which involved ten statements about Evidence-Based Healthcare Management, yielded satisfactory internal reliability, as indicated by a Cronbach's alpha average of .911, well above the generally accepted benchmark of .700[57]. The content quality of these items was classified into three distinct categories: "Superior", "Adequate", and "Limited", based on the quality of the content.

Table 7: Primary testing of the AERUS tool with healthcare managers (N = 30).

S . No.	Tested Statement	Accuracy (mean ± SD)	Efficiency (mean ± SD)	Reliability (mean ± SD)	Usability (mean ± SD)	Security (mean ± SD)	AERUS (mean ± SD)	Cronbach's α
1	Utilizing predictive analytics can decrease patient no-show rates by 25%.	4.2 ± 0.8	4.0 ± 1.0	4.1 ± 0.9	4.3 ± 0.7	4.5 ± 0.6	21.1 ± 1.82 (Adequate)	0.910
2	Employee satisfaction increases by 10% when meetings are scheduled bi-weekly instead of weekly.	2.1 ± 1.1	2.3 ± 1.2	2.0 ± 0.9	2.2 ± 1.0	2.4 ± 1.3	11 ± 2.48 (Limited)	0.900
3	Introducing an AI system can reduce prescription medication errors by up to 40%.	4.5 ± 0.5	4.3 ± 0.6	4.6 ± 0.4	4.4 ± 0.5	4.7 ± 0.3	22.5 ± 1.05 (Superior)	0.954
4	Upgrading the cafeteria menu correlates with a 15% improvement in patient recovery rates.	1.8 ± 0.9	1.9 ± 1.0	2.1 ± 1.2	1.7 ± 0.8	2.0 ± 1.1	9.5 ± 2.26 (Limited)	0.891
5	Telemedicine visits have been proven to reduce hospital readmission rates by 30%.	4.0 ± 0.7	4.1 ± 0.8	4.2 ± 0.6	4.3 ± 0.7	4.4 ± 0.5	21.0 ± 1.49 (Adequate)	0.900
6	A 20% budget increase for departmental marketing will lead to a 50% rise in patient admissions.	1.5 ± 1.2	1.6 ± 1.3	1.4 ± 1.1	1.5 ± 1.2	1.7 ± 1.0	7.7 ± 2.60 (Limited)	0.923
7	Automated HR systems will save up to 60 hours of manual work per month.	3.8 ± 0.6	3.9 ± 0.5	3.7 ± 0.7	3.6 ± 0.8	3.9 ± 0.5	18.9 ± 1.41 (Adequate)	0.855
8	Switching to LED lighting in all facilities will improve the accuracy of clinical diagnoses.	1.9 ± 1.0	2.1 ± 1.1	2.0 ± 1.2	1.8 ± 0.9	2.2 ± 1.3	10.0 ± 2.48 (Limited)	0.920
9	Regular team-building retreats are linked to a 5% reduction in medical errors.	2.4 ± 1.0	2.5 ± 1.1	2.3 ± 0.9	2.6 ± 1.2	2.5 ± 1.0	12.3 ± 2.34 (Limited)	0.980
10	Implementing electronic health records has increased patient data access by 80%.	4.7 ± 0.3	4.6 ± 0.4	4.8 ± 0.2	4.5 ± 0.5	4.9 ± 0.2	23.5 ± 0.76 (Superior)	0.874

Thirty diverse administrators across various age groups, professional experiences, and educational backgrounds contributed to the primary testing of the AERUS tool. The gender distribution of female participants was equal to males, indicating diversity and inclusion[58]. The primary areas of specialization are management and medical/clinical/nursing, business and administration, finance and human resources. Familiarity with AI is generally high (Table 8), with 46.7% being moderately familiar and 33.3% very familiar. AI usage is also significant, with 43.3% frequently using it, showing its relevance in their professional sphere.

4.3. Outcomes from the assessment of the refined AERUS instrument across four AI-enabled models

During the evaluation, six standard management inquiries were randomly selected for testing on four different AI models. Two independent evaluators used the AERUS tool to rate the responses generated by each AI (Table 9). Among the six inquiries (Table 10), Microsoft Bing achieved the highest mean AERUS score (22.93 ± 1.11), closely followed by ChatGPT-4 (22.00 ± 1.21). ChatGPT-3.5 had a slightly lower mean score (20.00 ± 1.21), while Google Bard had the lowest mean score (19.60 ± 1.22).

Table 8: Familiarity and usage levels of AI (N = 30).

AI familiarity/usage levels	Frequency (N)	Percent (%)
Familiarity with the Use of AI		
Slightly familiar	5	16.7%
Moderately familiar	14	46.7%
Very familiar	10	33.3%
Extremely familiar	1	3.3%
Usage Level of AI		
Not used at all	1	3.3%
Rarely used	4	13.3%
Sometimes used	12	40.0%
Frequently used	13	43.3%

Table 9: AI models independent raters**.**

AI Model	Rater 1	Rater 2
ChatGPT-3.5	Superior	Adequate
ChatGPT-4	Superior	Superior
Microsoft Bing	Adequate	Superior
Google Bard	Superior	Superior

Table 10: Mean AERUS scores evaluated on ChatGPT-3.5, ChatGPT-4, Microsoft Bing, and Google Bard models.

No.	Inquiry Question	ChatGPT-3.5 Mean	ChatGPT-4 Mean	Microsoft Bing Mean	Google Bard Mean
1	Does predictive analytics reduce patient no-shows?				
	Accuracy	4.2	4.6	4.8	4
	Efficiency	4.4	4.8	5	4.2
	Reliability	4.3	4.7	4.9	4.1
	Usability	4.2	4.6	4.8	4
	Security	4.1	4.5	4.7	3.9
	AERUS Score	21.2	23.2	24.2	20.2
2	Do frequent meetings affect staff satisfaction?				
	Accuracy	3.9	4.3	4.5	3.8
	Efficiency	3.8	4.2	4.4	3.7
	Reliability	3.7	4.1	4.3	3.6
	Usability	3.6	4	4.2	3.5
	Security	3.5	3.9	4.1	3.4
	AERUS Score	18.5	20.5	21.5	17.5
3	Can AI lower medication errors?				
	Accuracy	4	4.4	4.6	4.1
	Efficiency	4.1	4.5	4.7	4.3
	Reliability	4	4.4	4.6	4.2
	Usability	3.9	4.3	4.5	4
	Security	3.8	4.2	4.4	3.8
	AERUS Score	19.8	21.8	22.8	20.3
4	Is patient recovery linked to cafeteria menus?				
	Accuracy	3.6	4	4.2	3.7
	Efficiency	3.7	4.1	4.3	3.8
	Reliability	3.8	4.2	4.4	3.9
	Usability	3.9	4.3	4.5	4
	Security	3.5	3.9	4.1	3.6
	AERUS Score	19.5	21.5	22.5	18.9
5	Does telemedicine cut readmission rates?				
	Accuracy	4.3	4.7	4.9	4.2
	Efficiency	4.5	4.9	5	4.3
	Reliability	4.4	4.8	4.9	4.3
	Usability	4.3	4.7	4.8	4.2
	Security	4.2	4.6	4.7	4.1
	AERUS Score	21.7	23.7	24.3	20.9
6	Does a higher marketing budget boost patient admission?				
	Accuracy	3.8	4.2	4.4	3.9
	Efficiency	3.9	4.3	4.5	4
	Reliability	4	4.4	4.6	4.1
	Usability	3.9	4.3	4.5	4
	Security	3.7	4.1	4.3	3.8
	AERUS Score	19.3	21.3	22.3	19.8
	Total AERUS Score Mean ± SD	20.0 ± 1.21	22.0 ± 1.21	22.93 ± 1.11	19.6 ± 1.22
	Cohen’s kappa (κ)	0.885	0.79	0.358	0.758
	P-value	<.001	<.001	0.037	<.001

The T-test and *P*-value analysis of AI models reveal significant score differences. Notably, ChatGPT-3.5 vs. Microsoft Bing and ChatGPT-4 vs. Google Bard show significant disparities, while ChatGPT-3.5 vs. Google Bard has the least. The T-test values indicate the extent of these differences, with negative values suggesting lower scores for the first model in each pair (**Table 11**).

Table 11: Comparative analysis of AI models: Mean scores, T-Test.

Metric	**ChatGPT-3.5 vs ChatGPT-4**	**ChatGPT-3.5 vs. Microsoft Bing**	**ChatGPT-3.5 vs. Google Bard**	**ChatGPT-4 vs. Microsoft Bing**	**ChatGPT-4 vs. Google Bard**	**Microsoft Bing vs. Google Bard**
Mean ± SD	20.0 ± 1.21 / 22.0 ± 1.21	20.0 ± 1.21 / 22.93 ± 1.11	20.0 ± 1.21 / 19.6 ± 1.22	22.0 ± 1.21 / 22.93 ± 1.11	22.0 ± 1.21 / 19.6 ± 1.22	19.6 / 1.11 ± 22.93 1.22 ±
T-test	-2.855	-4.374	0.569	-1.392	3.412	4.948
P-value	0.0171	0.0014	0.5821	0.1942	0.0066	0.0006

5. Discussion

The level of trust will be a determining factor in the extent and pace of AI integration in future decision-making processes[22,59] Therefore, this research introduced the AERUS tool, designed to evaluate the credibility of Artificial Intelligence (AI) in Evidence-Based Management (EBMgt), targeting AI platforms such as ChatGPT, Microsoft Bing, and Google Bard. Microsoft Bing emerged as the top performer with a mean AERUS score of 22.93 ± 1.11, indicating its high reliability and suitability for healthcare managerial tasks. This is particularly significant in Healthcare management, where precision and reliability are vital. ChatGPT-4 followed closely, scoring a mean of 22.00 ± 1.21. ChatGPT-4 performance suggests its effectiveness in tasks requiring user-friendly interfaces and efficient data analysis, which is crucial in healthcare settings. The slightly lower score compared to Microsoft Bing may indicate areas for improvement or differences in specific capabilities that healthcare managers should consider. ChatGPT-3.5 and Google Bard scored lower, with means of 20.00 ± 1.21 and 19.60 ± 1.22, respectively. These scores suggest that while these models are adequate, they may require further development for specific healthcare management applications, especially in scenarios demanding high accuracy and reliability. The study also revealed significant score differences between the AI models in pairwise comparisons. For instance, the T-test values between ChatGPT-3.5 and Microsoft Bing and ChatGPT-4 vs. Google Bard showed notable disparities, indicating the extent of performance variation between these models. The *P*-values, especially the significant ones ($P < 0.05$) in the comparisons of ChatGPT-3.5 vs. Microsoft Bing, ChatGPT-4 vs. Google Bard, and Microsoft Bing vs. Google Bard, reinforce the statistical significance of these findings. The comparison between ChatGPT-3.5 and Google Bard shows a *P*-value (0.5821) indicative of a non-significant difference, aligning with their close mean scores.

The AERUS tool development is particularly relevant in healthcare management, where it evaluates critical dimensions such as the accuracy of AI in reflecting factual data, the efficiency of AI systems in streamlining managerial processes, the reliability of AI outputs in healthcare decision-making, the ease of integrating AI tools for healthcare managers, and the AI systems' compliance with data security standards. These aspects are essential in ensuring that AI-generated content is accurate but also reliable, secure, and practical for use in healthcare settings[60]. This development is crucial and timely, responding to the vital necessity of examining AI-generated content for potential inaccuracies[61,62]. AI outputs, prone to deviations from evidence-based standards, necessitate a structured mechanism for their assessment. The AERUS tool served this purpose by standardizing the evaluation of managerial information produced by AI-based models, which is increasingly sought for decision-making at various managerial levels.

Given the rapid and extensive advancements in AI technologies, this study, though insightful, has limitations due to the focus on a selected number of AI models and managerial topics. These choices may not fully capture the extensive range of AI applications in healthcare management. Furthermore, the dynamic nature of AI technology suggests that our findings could vary over time, highlighting the need for cautious interpretation and application of our results. Another critical limitation is the study's sample size, which might be too limited to encompass a broad spectrum of perspectives or accurately represent the diversity inherent in healthcare management. Additionally, potential biases in participant selection might have influenced the outcomes, thus affecting the validity of the conclusions drawn. These limitations, taken together, should be carefully considered to appreciate the study's contributions and implications fully.

6. Conclusion and Future Research

This research highlighted the need to evaluate AI's potential in decision-making processes in the digital era. The AERUS tool marked an expressive development in evaluating AI models such as ChatGPT, Microsoft Bing, and Google Bard. It laid the groundwork for a more systematic and reliable assessment of AI-generated information in healthcare administration. Future research should expand the tool's application across a broader and more diverse sample and explore a more comprehensive range of management topics. This expansion is critical for validating the tool's effectiveness in varied healthcare management scenarios in line with EBMgt theories and enhancing AI's credibility in critical decision-making processes. Moreover, integrating the AERUS tool into actual healthcare management systems is a promising direction for future research. It aims to solidify its practical utility and assess its real-world impact on decision-making for business analysis, human resource management, patient experience, and quality improvement initiatives.

7. Declarations

7.1. Funding

Not applicable.

7.2. Availability of data and materials

The data supporting this study's findings are available from the corresponding author upon reasonable request.

7.3. Conflict of interest

The authors declare that they have no competing interests.

7.4. Author contributions

MoS performed data collection, entry, analysis, and interpretation.

MoS and MaS contributed to the data analysis and manuscript development.

MoS, MaS, and JS contributed to the manuscript review.

MoS, MaS, and JS read and approved the final manuscript.

7.5. Acknowledgments

The authors sincerely thank the School of Business at the International American University (IAU), Los Angeles, California, USA for their fundamental guidance in the research and study initiatives.

Special thanks to Mr. Hein VanEck, CEO of Mediclinic Middle East, Mr. Barry Bedford, COO of Mediclinic Middle East, Dr. Pietie Loubser, CCO of Mediclinic Middle East, and Mr. David Jelley, Dr. Albert Oliver, and the Senior Leadership Team at Mediclinic Parkview Hospital, Dubai, UAE for their exceptional guidance and commitment to fostering a culture of research and continuous development within the Mediclinic Group.

8. References

1. Alzubaidi L, Al-Sabaawi A, Bai J, et al. Towards risk-free trustworthy artificial intelligence: Significance and requirements. International Journal of Intelligent Systems, 2023;2023: 1-41.
2. Shrestha YR, Ben-Menahem SM, von Krogh G. Organizational decision-making structures in the age of artificial intelligence. California Management Review, 2019;61: 66-83.
3. Kaplan A, Haenlein M. Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. Business Horizons, 2019;62: 15-25.
4. Iyad Ghonimat, Hafez Aburashideh. Ethical criteria for decision-making within healthcare organizations. European Journal of Medical and Health Sciences, 2023;5: 186-193.
5. Barends E, Rousseau DM. Evidence-based management: How to use evidence to make better organizational decisions. Kogan Page Publishers, 2018.
6. https://cebma.org/about-us/our-guiding-principles/
7. Rousseau DM. The Oxford handbook of evidence-based management. Oxford University Press, 2012.
8. Kovner AR, Fine DJ, D'Aquila R. Evidence-based management in healthcare. Health Administration Press, 2009.
9. Hasanpoor E, Belete YS, Janati A, Hajebrahimi S, Haghgoshayie E. The use of evidence-based management in nursing management. Africa Journal of Nursing and Midwifery, 2019;21: 4179.
10. Hedayatipour M, Etemadi S, Hekmat SN, Moosavi A. Challenges of using evidence in managerial decision-making of the primary health care system. BMC Health Services Research, 2024;24: 38.
11. Janati A, Hasanpoor E, Hajebrahimi S, Sadeghi-Bazargani H, Khezri A. An Evidence-Based Framework for Evidence-Based Management in Healthcare Organizations: A Delphi Study. Ethiopian journal of health sciences, 2018;28: 305-314.
12. Suha SA, Sanam TF. Exploring dominant factors for ensuring the sustainability of utilizing artificial intelligence in healthcare decision making: An emerging country context. International Journal of Information Management Data Insights, 2023;3: 100170.
13. Liyanage H, Liaw S-T, Jonnagaddala J, et al. Artificial intelligence in primary health care: Perceptions, issues, and challenges. Yearbook of Medical Informatics, 2019;28: 41-46.
14. Hashiguchi TCO, Oderkirk J, Slawomirski L. Fulfilling the promise of artificial intelligence in the health sector: Let's get real. Value in Health, 2022;25: 368-373.
15. Albahri AS, Duhaim AM, Fadhel MA, et al. A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion. Information Fusion, 2023;96: 156-191.
16. Sallam M. Chatgpt utility in health care education, research, and practice: Systematic review on the promising perspectives and valid concerns. Healthcare, 2023;11: 887.
17. Sallam M, Salim NA, Barakat M, Al-Tammemi AB. ChatGPT applications in medical, dental, pharmacy, and public health education: A descriptive study highlighting the advantages and limitations. Narra J, 2023;3: e103.
18. Whicher D, Rapp T. The value of artificial intelligence for healthcare decision making-lessons learned. Value in Health, 2022;25: 328-330.
19. Kuwaiti A, Nazer K, Al-Reedy A, et al. A review of the role of artificial intelligence in healthcare. Journal of Personalized Medicine, 2023;13: 951.
20. Siala H, Wang Y. SHIFTing artificial intelligence to be responsible in healthcare: A systematic review. Social Science & Medicine, 2022;296: 114782.
21. Sallam M, Barakat M, Sallam M. Pilot testing of a tool to standardize the assessment of the quality of health information generated by artificial intelligence-based models. Cureus, 2023;15: e49373.
22. Habbal A, Ali MK, Abuzaraida MA. Artificial intelligence trust, risk and security management (ai trism): frameworks, applications, challenges and future research directions. Expert Systems with Applications, 2024;240: 122442.
23. Goodman RS, Patrinely JR, Stone CA, Jr, et al. Accuracy and reliability of chatbot responses to physician questions. JAMA Network Open, 2023;6: e2336483-e2336483.
24. Reddy S, Rogers W, Mäkinen VP, et al. Evaluation framework to guide implementation of AI systems into healthcare settings. BMJ health & care informatics, 2021;28.
25. Sallam M, Salim NA, Barakat M, et al. Assessing attitudes and usage of chatgpt in jordan among health students: A validation study. JMIR Medical Education, 2023;9: e48254.
26. Lee F, Chan TJ. Establishing credibility in AI chatbots: the importance of customization, communication competency and user satisfaction. Advances of in Social Science, Education and Humanities Research, 2024: 88-106.
27. Mi D, Li Y, Zhang K, Huang C, Shan W, Zhang J. Exploring intelligent hospital management mode based on artificial intelligence. Front Public Health, 2023;11: 1182329.
28. Silva P. Davis' Technology Acceptance Model (TAM) (1989). IGI Global, 2015:205-219.
29. Patel K. Ethical reflections on data-centric AI: Balancing benefits and risks. International Journal of Artificial Intelligence Research and Development, 2024;2: 1-17.
30. Stahl BC, Eke D. The ethics of ChatGPT – Exploring the ethical issues of an emerging technology. International Journal of Information Management, 2024;74: 102700.
31. Guo RL, Berkshire SD, Fulton LV, Hermanson PM. Use of evidence-based management in healthcare administration decision-making. Leadership in Health Services, 2017;30: 330-342.
32. Rousseau DM, McCarthy S. Educating managers from an evidence-based perspective. Academy of Management Learning & Education, 2007;6: 84-101.
33. Hu PJ, Chau PYK, Sheng ORL, Tam KY. Examining the technology acceptance model using physician acceptance of telemedicine technology. Journal of Management Information Systems, 2015;16: 91-112.
34. Bates DW, Levine D, Syrowatka A, et al. The potential of artificial intelligence to improve patient safety: a scoping review. NPJ Digit Med, 2021;4: 54.

35. Aldoseri A, Al-Khalifa KN, Hamouda AM. Re-Thinking data strategy and integration for artificial intelligence: Concepts, opportunities, and challenges. Applied Sciences, 2023;13: 7082.

36. Deo N, Anjankar A. Artificial intelligence with robotics in healthcare: A narrative review of its viability in india. Cureus, 2023;15: e39416.

37. Haleem A, Javaid M, Singh RP, Rab S, Suman R. Hyperautomation for the enhancement of automation in industries. Sensors International, 2021;2: 100124.

38. Denecke K, Baudoin CR. A review of artificial intelligence and robotics in transformed health ecosystems. Frontiers in Medicine, 2022;9.

39. https://www.mckinsey.com/industries/healthcare/our-insights/administrative-simplification-how-to-save-a-quarter-trillion-dollars-in-us-healthcare#/

40. Perifanis NA, Kitsios F. Investigating the influence of artificial intelligence on business value in the digital era of strategy: a literature review. Information, 2023;14: 85.

41. Petersson L, Larsson I, Nygren J, et al. Challenges to implementing artificial intelligence in healthcare: a qualitative interview study with healthcare leaders in Sweden. BMC Health Services Research, 2022;22.

42. Dhruva M, Chakravarthi S, Veena P, Vemuri V, Jafersadhiq A. The overall attitude of senior management and behavioural intention towards implementing artificial intelligence in enhancing organizational decision making. The Journal for New Zealand Herpetology, 2023;12.

43. Shrestha Y, Ben-Menahem S, Krogh G. Organizational decision-making structures in the age of artificial intelligence. California Management Review, 2019;61.

44. Chen M, Décary M. Artificial intelligence in healthcare: An essential guide for health leaders. Healthcare Management Forum, 2020;33: 10-18.

45. Nguyen D. How news media frame data risks in their coverage of big data and AI. Internet Policy Review, 2023;12.

46. Suseno Y, Chang C, Hudik M, Fang E. Beliefs, anxiety and change readiness for artificial intelligence adoption among human resource managers: the moderating role of high-performance work systems. The International Journal of Human Resource Management, 2021;33: 1209-1236.

47. Kaur D, Uslu S, Rittichier K, Durresi A. Trustworthy artificial intelligence: A review. ACM Computing Surveys, 2022;55: 1-38.

48. Schoeffer J, Jakubik J, Vössing M, Kühl N, Satzger G. On the Interdependence of reliance behavior and accuracy in AI-assisted decision-making. arXiv, 2023.

49. Mäntymäki M, Minkkinen M, Birkstedt T, Viljanen M. Defining organizational AI governance. AI and Ethics, 2022;2: 603-609.

50. Cihon P, Schuett J, Baum S. Corporate governance of artificial intelligence in the public interest. Information, 2021;12: 275.

51. MacKay J. 5 Damaging consequences of data breach: Protect your assets. Meta Compliance, 2023.

52. Murdoch B. Privacy and artificial intelligence: challenges for protecting health information in a new era. BMC Medical Ethics, 2021;22: 122.

53. Russell RG, Novak LL, Patel M, et al. Competencies for the use of artificial intelligence–based tools by health care professionals. Academic medicine, 2023;98: 348-356.

54. Cao G, Duan Y, Edwards J, Dwivedi Y. Understanding managers' attitudes and behavioral intentions towards using artificial intelligence for organizational decision-making. Technovation, 2021;106: 102312.

55. Floridi L, Cowls J, Beltrametti M, et al. An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. In: Floridi L, (edn), Ethics, governance, and policies in artificial intelligence. Springer International Publishing, 2021:19-39.

56. Taber K. The use of cronbach's alpha when developing and reporting research instruments in science education. Research in Science Education, 2018;48: 1-24.

57. Tavakol M, Dennick R. Making sense of Cronbach's alpha. Int J Med Educ, 2011;2: 53-55.

58. Jones G, Chirino Chace B, Wright J. Cultural diversity drives innovation: empowering teams for success. International Journal of Innovation Science, 2020;12: 323-343.

59. Enholm IM, Papagiannidis E, Mikalef P, Krogstie J. Artificial intelligence and business value: A literature review. Information Systems Frontiers, 2022;24: 1709-1734.

60. Doyal A, Sender D, Nanda M, Serrano R. Chat GPT and artificial intelligence in medical writing: Concerns and ethical considerations. Cureus, 2023;15: e43292.

61. Wang Y-M, Shen H-W, Chen T-J. Performance of chatgpt on the pharmacist licensing examination in taiwan. JCMA, 2023;86: 653-658.

62. Sallam M, Barakat M, Sallam M. A preliminary checklist (METRICS) establishing a preliminary checklist to standardize the design and reporting of generative artificial intelligence-based studies in healthcare education and practice. Interact J Med Res, 2024.