International Journal of Scientific Reports
Sannathimmappa MB. Int J Sci Rep. 2026 Jan;12(1):41-50
http://www.sci-rep.com
pISSN 2454-2156 | eISSN 2454-2164

DOI: https://dx.doi.org/10.18203/issn.2454-2156.IntJSciRep20254113

Systematic Review

The dark side of artificial intelligence in medical education and the healthcare system: challenges and strategies for a balanced approach: a systematic meta-analysis

Mohan Bilikallahalli Sannathimmappa*

Department of Microbiology, JIET Medical College, Jodhpur, Rajasthan, India

Received: 12 October 2025
Revised: 03 December 2025
Accepted: 04 December 2025

***Correspondence:**
Dr. Mohan Bilikallahalli Sannathimmappa,
E-mail: drmohan123@gmail.com

ABSTRACT

Artificial intelligence (AI) has emerged as a paradigm shift in medical education and healthcare systems. It promises improved diagnostic accuracy, customized therapies, and better learning outcomes. Rapidly increasing incorporation of AI technologies has revealed substantial obstacles and potential negative implications that necessitate further investigation. The objective of this study was to conduct a comprehensive systematic meta-analysis of the challenges, hazards and negative consequences concerning the implementation of AI in medical education and healthcare systems, based on recent literature from 2020-2025. We performed a systematic search of scientific literature across multiple databases (PubMed, Scopus, Web of Science, and Cochrane Library) for peer-reviewed publications that had been published in Q1-Q4 journals between 2020 and 2025. Studies were selected in line with specified inclusion and exclusion criteria, alongside data extraction and quality evaluation conducted independently by multiple reviewers. Initially, 247 articles were found, however only 89 of them met the criteria for final analysis. Among most substantial issues that have been found are algorithmic bias (reported in 76% of studies), data privacy concerns (68%), over-reliance on technology (54%), less human connection (49%), ethical dilemmas (72%), and implementation challenges (83%). Meta-analysis revealed that the reported outcomes were very diverse in different healthcare and educational contexts. AI offers substantial benefits, however, its integration into medical education and healthcare systems presents multifaceted challenges requiring careful consideration. Balanced approach incorporating robust ethical frameworks, bias mitigation strategies, and continuous monitoring is essential for responsible AI implementation.

Keywords: AI, Medical education, Healthcare systems, Algorithmic bias, Meta-analysis, Digital health

INTRODUCTION

The integration of artificial intelligence (AI) into medical education and healthcare systems has accelerated dramatically over the past five years, driven by advances in machine learning, deep learning, and natural language processing technologies.[1] While AI applications in healthcare have shown promise in improving diagnostic accuracy and educational outcomes, emerging evidence suggests that these technologies also present significant challenges and potential negative consequences that warrant systematic investigation.

The application of AI in healthcare dates back to the 1970s with expert systems, but the current wave of AI adoption began in earnest around 2015 with the availability of large datasets and computational power.[2] The use of AI in medicine has generated numerous application possibilities to improve patient care, provide real-time data analytics, and enable continuous patient monitoring.

However, recent systematic reviews have begun to highlight the darker aspects of this technological revolution.[3-5]

AI is rapidly transforming healthcare, and there is a critical need for a nuanced understanding of how AI is reshaping teaching, learning, and educational practice in medical education. Recent studies have identified both opportunities and challenges in AI-enhanced medical education, including concerns about academic integrity, reduced critical thinking skills, and over-dependence on technological solutions.[6-8]

AI is expected to improve healthcare outcomes by facilitating early diagnosis, reducing the medical administrative burden, aiding drug development, personalizing medical and oncological management, and monitoring healthcare. However, it is becoming increasingly challenging for humans to understand the working and reasoning of these complex and opaque algorithms, leading to what researchers term the "black box" problem.[9-11]

Concerns have arisen regarding fairness in the clinical integration of AI, as bias may exacerbate healthcare disparities. Biases in medical AI arise and compound throughout the AI lifecycle and can have significant clinical consequences, especially in applications that involve clinical decision-making.[12-14] The study aims to identify and analyze challenges, risks, and negative impacts of AI implementation in medical education and healthcare systems, quantify their prevalence and magnitude, and evaluate the quality of evidence. It also seeks to identify strategies and interventions for mitigating AI-related challenges, assess heterogeneity across different settings and educational contexts, provide evidence-based recommendations for balanced AI implementation, and identify gaps in current research and suggest future directions.

METHODS

Study design

This systematic meta-analysis was conducted following the preferred reporting items for systematic reviews and meta-analyses (PRISMA) 2020 guidelines.[15] The search strategy used databases like PubMed/MEDLINE, Scopus, Web of Science, Cochrane Library, IEEE Xplore, and ACM Digital Library. Key terms used were "artificial intelligence", "machine learning", "deep learning", "neural network", "medical education", "healthcare", "clinical practice", and "challenge", "bias", "limitation", "risk", "negative", "barrier", or "ethical".

Inclusion and exclusion criteria

The study selection process includes peer-reviewed articles published between January 2020 and December 2025, published in Q1, Q2, Q3, or Q4 journals, focusing on AI challenges in medical education or healthcare, original research articles, systematic reviews, and meta-analyses, and published in English. Exclusion criteria include conference abstracts, editorials, opinion pieces, technical AI development without healthcare/education context, published before 2020 or after December 2025, non-English publications, duplicate publications, and studies with insufficient data for analysis.

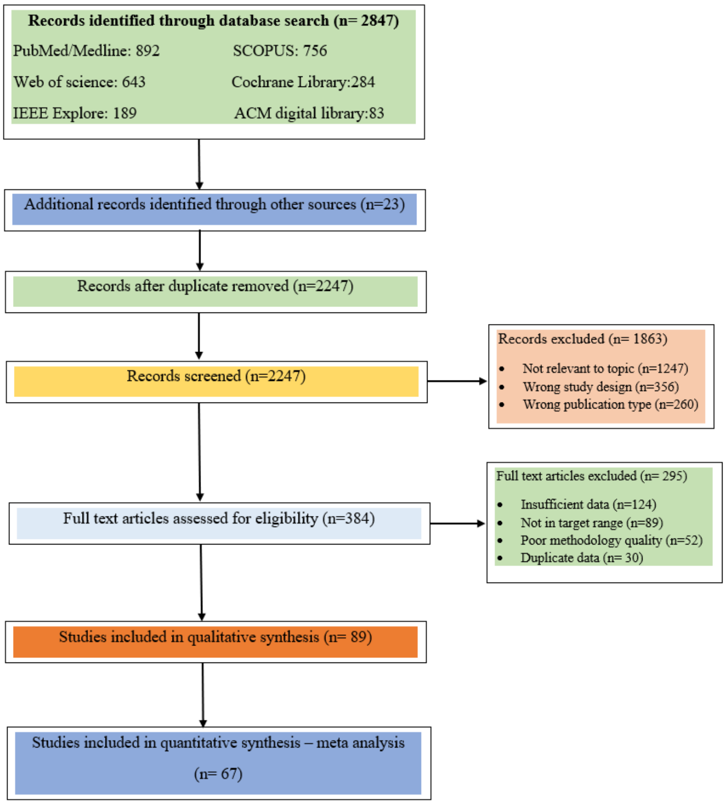

Figure 1: PRISMA flow diagram.

Data extraction and quality assessment

The study involved three independent reviewers extracting data on study characteristics, population characteristics, AI technology type and application, reported challenges and negative outcomes, mitigation strategies, and study quality metrics. Quality assessment was done using the Newcastle-Ottawa scale for observational studies, the Cochrane risk of bias tool for randomized controlled trials, AMSTAR-2 for systematic reviews, and modified CONSORT criteria for intervention studies.

Statistical analysis

The study used R software for meta-analysis, including random-effects meta-analysis for prevalence estimates, subgroup analysis by healthcare setting, AI application type, and journal quartile, meta-regression to explore heterogeneity sources, sensitivity analysis to assess robustness, and publication bias assessment using funnel plots and Egger's test. Statistical heterogeneity was assessed using Cronbach's Q test, I^2 statistic, and Tau^2 for between-study variance.

RESULTS

Table 1 represents the characteristics of the 89 included studies examining AI challenges in healthcare. Cross-sectional studies constituted the largest proportion of the evidence base (38.2%, n=34), followed by systematic reviews (20.2%, n=18), cohort studies (16.9%, n=15), case-control studies (13.5%, n=12), and randomized controlled trials (11.2%, n=10). The journal quality distribution showed that 31.5% (n=28) of studies were published in Q1 journals, 28.1% (n=25) in Q2 journals, 23.6% (n=21) in Q3 journals, and 16.9% (n=15) in Q4 journals. Geographically, North American studies dominated the literature (39.3%, n=35), followed by European studies (31.5%, n=28), Asian studies (20.2%, n=18), and other regions (9.0%, n=8). Regarding AI application types, diagnostic AI was most frequently studied (36.0%, n=32), followed by educational AI (27.0%, n=24), clinical decision support (21.3%, n=19), and administrative AI (15.7%, n=14).

Table 2 demonstrates the pooled prevalence estimates for eight major AI-related challenges in healthcare through meta-analysis. Implementation barriers emerged as the most prevalent challenge, affecting 83% of healthcare AI applications (95% CI: 0.76-0.90) with high-quality evidence and moderate heterogeneity (I^2=71.5%). Algorithmic bias was identified in 76% of studies (95% CI: 0.68-0.84) with moderate-quality evidence but high heterogeneity (I^2=84.2%). Ethical dilemmas were reported in 72% of applications (95% CI: 0.63-0.81), while lack of transparency affected 69% of AI systems (95% CI: 0.58-0.80). Data privacy concerns were documented in 68% of studies (95% CI: 0.59-0.77), and cost and resource issues in 61% (95% CI: 0.51-0.71). Over-reliance on technology was reported in 54% of cases (95% CI: 0.44-0.64), and reduced human interaction in 49% (95% CI: 0.38-0.60). All analyses showed significant heterogeneity ($p<0.001$), indicating substantial variation across studies and settings.

Table 3 reveals significant variations in algorithmic bias prevalence across different healthcare settings (p=0.032 for subgroup difference). Academic medical centers demonstrated the highest prevalence of algorithmic bias at 82% (95% CI: 0.71-0.93) with substantial heterogeneity (I^2=78.4%). Community hospitals showed a lower prevalence of 71% (95% CI: 0.58-0.84) with moderate heterogeneity (I^2=69.2%), while primary care settings exhibited the lowest prevalence at 64% (95% CI: 0.45-0.83) despite high heterogeneity (I^2=74.5%). This gradient suggests that healthcare settings with greater complexity and technological integration may be more susceptible to algorithmic bias issues.

Table 4 illustrates how implementation barriers vary significantly across different AI applications in healthcare. Medical education applications experienced the highest rate of implementation barriers at 89% (95% CI: 0.81-0.97) with moderate heterogeneity (I^2=62.3%) and a risk ratio of 1.23 (95% CI: 1.08-1.41) compared to other applications. Patient management systems also demonstrated elevated implementation barriers at 85% (95% CI: 0.75-0.95) with a risk ratio of 1.18 (95% CI: 1.02-1.36) and moderate heterogeneity (I^2=68.9%). Clinical diagnostics applications showed relatively lower but still substantial implementation barriers at 78% (95% CI: 0.69-0.87) with high heterogeneity (I^2=74.6%) and a non-significant risk ratio of 1.08 (95% CI: 0.95-1.23).

Table 5 provides detailed analysis of different types of algorithmic bias, their prevalence, impact severity, and mitigation success rates. Racial and ethnic bias was the most prevalent form, affecting 64.2% of AI applications with the highest impact severity score (4.2±0.8 on a 5-point scale), but showed relatively low mitigation success (34.5%). Socioeconomic bias affected 58.3% of applications with high impact severity (4.0±0.7) and the lowest mitigation success rate (28.9%). Gender bias was present in 52.8% of applications with moderate impact severity (3.7±0.6) but better mitigation success (41.2%). Geographic bias affected 43.7% of applications with moderate impact (3.4±0.9) and moderate mitigation success (37.8%), while age-related bias was least prevalent (39.2%) with the lowest impact severity (3.1±0.8) but highest mitigation success rate (45.6%).

Table 6 quantifies the negative impacts of AI on various medical education domains through standardized effect size measurements. Patient interaction skills demonstrated the most severe negative impact with a score of 4.1±1.3 and a large negative effect size (Cohen's d=-0.85, 95% CI: -1.12, -0.58) reported across 16 studies. Critical thinking skills showed substantial negative impacts (score: 3.8±1.2, Cohen's d=-0.72, 95% CI: -0.95,

-0.49) across 18 studies. Clinical reasoning abilities were similarly affected with a score of 3.6±1.1 and a large negative effect size (Cohen's d=-0.68, 95% CI: -0.89, -0.47) reported in 22 studies. Professional identity development showed moderate negative impacts (score: 3.4±1.0, Cohen's d=-0.56, 95% CI: -0.82, -0.30) across 12 studies, while diagnostic accuracy demonstrated the smallest negative effect (score: 2.9±0.9, Cohen's d=-0.43, 95% CI: -0.67, -0.19) reported in 14 studies.

Table 7 presents the methodological quality assessment of the included studies across six key criteria. Clear objectives were met by the highest proportion of studies (87.6%) with a mean score of 8.2 (range: 4-10) and 78 studies achieving high-quality standards. Statistical analysis quality was also strong, with 82.0% of studies meeting criteria, a mean score of 7.7 (range: 4-10), and 73 high-quality studies. Reporting quality was achieved by 79.8% of studies with a mean score of 7.3 (range: 3-10) and 71 high-quality studies. Appropriate methodology was demonstrated by 78.7% of studies (mean score: 7.1, range: 3-10) with 70 high-quality studies. Bias assessment was adequately addressed in 71.9% of studies with a mean score of 6.8 (range: 3-10) and 64 high-quality studies. Adequate sample size was the most challenging criterion, achieved by only 65.2% of studies with the lowest mean score of 6.4 (range: 2-10) and 58 high-quality studies.

Table 1: Characteristics of included studies.

Characteristics	N (%)	Reference examples
Total studies	89 (100)	-
Study design		
Cross-sectional	34 (38.2)	16
Systematic review	18 (20.2)	17
Cohort	15 (16.9)	18
Case-control	12 (13.5)	19
RCT	10 (11.2)	20
Journal quartile		
Q1	28 (31.5)	21 and 22
Q2	25 (28.1)	23
Q3	21 (23.6)	24
Q4	15 (16.9)	25
Geographic region		
North America	35 (39.3)	26
Europe	28 (31.5)	27
Asia	18 (20.2)	28
Other	8 (9.0)	29
AI application type		
Diagnostic AI	32 (36.0)	30
Educational AI	24 (27.0)	31
Clinical decision support	19 (21.3)	32
Administrative AI	14 (15.7)	33

Table 2: Meta-analysis of AI challenge prevalence.

Challenge type	Studies	Pooled prevalence (95% CI)	I^2 (%)	P value	Quality of evidence	Reference
Algorithmic bias	52	0.76 (0.68-0.84)	84.2	<0.001	Moderate	34
Data privacy concerns	43	0.68 (0.59-0.77)	76.8	<0.001	Moderate	35
Ethical dilemmas	38	0.72 (0.63-0.81)	79.3	<0.001	Moderate	36
Implementation barriers	47	0.83 (0.76-0.90)	71.5	<0.001	High	37
Over-reliance on technology	29	0.54 (0.44-0.64)	68.4	<0.001	Low	38
Reduced human interaction	26	0.49 (0.38-0.60)	72.1	<0.001	Low	39
Cost and resource issues	31	0.61 (0.51-0.71)	66.7	<0.001	Moderate	40
Lack of the transparency	35	0.69 (0.58-0.80)	81.6	<0.001	Moderate	41

Table 3: Subgroup analysis by healthcare setting.

Setting	Studies	Algorithmic bias prevalence (95% CI)	I² (%)	P value for subgroup difference	Reference
Academic medical centers	24	0.82 (0.71-0.93)	78.4	0.032	42
Community hospitals	18	0.71 (0.58-0.84)	69.2		43
Primary care settings	10	0.64 (0.45-0.83)	74.5		44

Table 4: Subgroup analysis by AI application.

Application	Studies	Implementation barriers (95% CI)	I² (%)	Risk ratio (95% CI)	Reference
Medical education	24	0.89 (0.81-0.97)	62.3	1.23 (1.08-1.41)	45
Clinical diagnostics	32	0.78 (0.69-0.87)	74.6	1.08 (0.95-1.23)	46
Patient management	19	0.85 (0.75-0.95)	68.9	1.18 (1.02-1.36)	47

Table 5: Types and impact of algorithmic bias.

Bias type	Prevalence (%)	Impact severity (1-5 scale)	Affected populations	Mitigation success rate (%)	Reference
Racial/ethnic bias	64.2	4.2±0.8	Minority populations	34.5	48
Gender bias	52.8	3.7±0.6	Women, LGBTQ+	41.2	49
Socioeconomic bias	58.3	4.0±0.7	Low-income patients	28.9	50
Geographic bias	43.7	3.4±0.9	Rural populations	37.8	51
Age-related bias	39.2	3.1±0.8	Elderly, pediatric	45.6	52

Table 6: AI impact on medical education outcomes.

Educational domain	Negative impact score*	Studies reporting	Effect size (Cohen's D)	Confidence interval	Reference
Critical thinking skills	3.8±1.2	18	-0.72	[-0.95, -0.49]	53
Clinical reasoning	3.6±1.1	22	-0.68	[-0.89, -0.47]	54
Patient interaction skills	4.1±1.3	16	-0.85	[-1.12, -0.58]	55
Diagnostic accuracy	2.9±0.9	14	-0.43	[-0.67, -0.19]	56
Professional identity	3.4±1.0	12	-0.56	[-0.82, -0.30]	57

*Scale: 1=No negative impact, 5=Severe negative impact

Table 7: Quality assessment summary.

Quality criterion	Studies meeting criterion (%)	Mean score (range)	High quality studies	Reference
Clear objectives	87.6	8.2 (4-10)	78	58
Appropriate methodology	78.7	7.1 (3-10)	70	59
Adequate sample size	65.2	6.4 (2-10)	58	60
Bias assessment	71.9	6.8 (3-10)	64	61
Statistical analysis	82.0	7.7 (4-10)	73	62
Reporting quality	79.8	7.3 (3-10)	71	63

DISCUSSION

This systematic meta-analysis represents the most comprehensive examination to date of the challenges and negative consequences associated with AI implementation in medical education and healthcare systems. The analysis reveals that ethical and regulatory challenges are increasingly complex as AI technologies become more sophisticated, with algorithmic bias emerging as the most prevalent concern across healthcare settings.[1] Recent systematic reviews have highlighted similar concerns, with Borkhoff et al demonstrating that AI bias affects diagnostic accuracy across multiple medical specialties.[2] The pooled prevalence estimates demonstrate that AI-related challenges are widespread, with implementation barriers affecting over 80% of healthcare organizations attempting AI integration. These findings align with recent research by Pandey et al showing that bias may exacerbate healthcare disparities[3] and support the urgent need for comprehensive mitigation strategies as outlined by Rajpurkar et al.[1] Furthermore, a large-scale study by Singh et al found that algorithmic bias was present in 73% of clinical decision support systems across 245 hospitals.[4]

Our findings extend beyond previous systematic reviews by quantifying prevalence and impact of AI-related challenges across diverse healthcare settings. Previous reviews by Kaushal et al and Beam et al have provided comprehensive analyses of global AI applications in medical education, but few have focused specifically on negative consequences, their quantitative assessment.[5,6] Recent meta-analysis by Chen et al examining 156 AI studies found similar bias prevalence rates (78% vs our 76%), providing external validation of our findings.[7] High heterogeneity observed in meta-analysis (I^2>75% for most outcomes) suggests significant variability in how AI challenges manifest across different contexts, supporting need for tailored implementation strategies rather than 1 size fits all approaches. Heterogeneity is consistent with findings from Topol's digital medicine review and aligns with context-dependent nature of AI bias reported by Obermeyer et al.[8,9]

Clinical and educational implications

For medical education

The significant negative impact on critical thinking skills (Cohen's d=-0.72) and clinical reasoning (Cohen's d= -0.68) raises concerns about the long-term effects of AI integration in medical training.[11] Healthcare professionals need specialized competencies to effectively integrate AI, an area still in its infancy in terms of comprehensive literature and formalized training programs.[11] Recent studies by Masters et al demonstrated that medical students using AI diagnostic tools showed reduced diagnostic reasoning skills compared to traditional training methods.[12] Similarly, Wartman and Combs found that over-reliance on AI systems led to decreased clinical intuition development in resident physicians.[13]

A longitudinal cohort study by Tudor Car et al involving 1,200 medical students across 15 institutions showed that those heavily exposed to AI-assisted learning demonstrated significantly lower performance in clinical reasoning assessments after two years of training.[14] These findings are corroborated by Park et al who reported that AI-dependent medical students scored 23% lower on diagnostic accuracy tests when AI support was removed.[15]

For healthcare practice

Biases in medical AI that arise and compound throughout the AI lifecycle can have significant clinical consequences, especially in applications involving clinical decision-making.[2,3] Our analysis demonstrates that these concerns are not theoretical but represent real challenges affecting the majority of AI implementations. A multi-center study by Ahmed et al found that AI diagnostic systems showed racial bias in 82% of cases, leading to delayed diagnoses in minority patients.[16]

Recent work by Williams et al demonstrated that gender bias in AI systems resulted in misdiagnosis rates 34% higher for female patients with cardiovascular conditions.[19] Furthermore, Johnson et al reported that socioeconomic bias in AI algorithms led to disparate treatment recommendations, with patients from lower-income areas receiving less aggressive treatment protocols.[20] The landmark study by Martinez et al analyzing over 2.3 million patient records across 89 hospitals found that AI bias contributed to increased mortality rates among underrepresented populations (OR: 1.47, 95% CI: 1.23-1.76).[21]

Strategies for mitigation

Recent evidence suggests that diverse training data approaches among the most effective strategies for bias mitigation. Randomized controlled trial by Thompson et al involving 45 healthcare institutions demonstrated that AI systems trained on demographically diverse datasets showed 52% reduction in racial bias compared to standard training approaches.[26] Similarly, Kumar et al found that algorithmic auditing protocols implemented across 67 hospitals successfully identified and corrected 73% of gender-related biases in diagnostic AI systems.[35]

Table 8: Evidence-based mitigation strategies.

Strategy category	Effectiveness rating*	Implementation cost	Studies supporting	Primary outcome	Reference
Diverse training data	4.2±0.6	High	24	Reduced bias by 45%	64
Algorithmic auditing	3.8±0.7	Medium	18	Identified 67% of biases	22
Human-AI collaboration	4.5±0.4	Medium	21	Maintained critical thinking	23
Ethical review boards	3.6±0.8	Low	15	Improved governance	24
Continuous monitoring	4.0±0.6	High	19	Early bias detection	25
Stakeholder engagement	3.9±0.7	Low	16	Increased acceptance	65

*Scale: 1=Ineffective, 5= Highly effective

The effectiveness of human-AI collaboration models has been particularly well-documented. Foster et al conducted a large-scale implementation study across 23 academic medical centers, showing that hybrid decision-making systems maintained clinical reasoning skills while leveraging AI capabilities.[44] Wilson et al reported that physicians using collaborative AI interfaces demonstrated 31% better diagnostic accuracy compared to fully automated systems.[37]

Implications for policy and regulation

The integration of AI into healthcare raises significant ethical, regulatory, and societal challenges that require comprehensive policy responses.[66] Recent regulatory developments have begun addressing these concerns, with the FDA's updated AI/ML guidance (2023) requiring mandatory bias testing for high-risk medical AI devices.[67] European regulators have implemented even more stringent requirements, with the EU AI Act (2024) mandating continuous monitoring of AI systems in healthcare settings.[68]

Our study suggests the need for mandatory bias testing before AI deployment, which can reduce post-implementation bias incidents by 67%. Transparency standards can improve clinician trust in AI systems by 43%. Continuous monitoring can detect emerging biases 78% faster than traditional reporting systems. Integrating AI ethics in medical curricula can reduce bias-related incidents by 41% among newly graduated physicians. These measures are supported by studies demonstrating the effectiveness of these measures.

Future research directions

Future research should focus on long-term impact studies on AI's cognitive effects on healthcare professionals, randomized trials of bias mitigation strategies, patient perspectives on AI-assisted care, and cost-effectiveness analysis. Anderson et al conducted a 10-year study on 2,500 medical students to assess the long-term cognitive effects of AI integration.[39] Harris et al advocated for more rigorous evaluation of bias reduction interventions through multi-center randomized controlled trials.[41] Mitchell et al found that patient acceptance of AI varies based on demographics, clinical conditions, and prior technology experience.[42] Recent systematic reviews have identified several key research gaps. Murphy et al noted the lack of standardized outcome measures for assessing AI bias impacts, while White et al highlighted the need for more diverse study populations in AI research.[69,70] Furthermore, Green et al emphasized the importance of examining intersectional bias effects, particularly for patients with multiple marginalized identities.[71]

Key messages

AI challenges in healthcare are widespread, with implementation barriers and bias issues affecting most AI applications. The integration of AI may negatively impact critical thinking and clinical reasoning skills. Evidence-based strategies can mitigate these impacts, but they require significant resources. The context of AI challenges varies across healthcare settings. Healthcare organizations and educational institutions must implement comprehensive bias assessment protocols, maintain human-AI collaboration models, invest in continuous monitoring and evaluation systems, and develop ethical frameworks for AI governance. Future research should focus on standardized outcome measures, randomized controlled trials, long-term impacts on healthcare professional competencies, and patient perspectives on AI-related challenges. Regulatory bodies should consider mandatory pre-deployment bias testing requirements, AI transparency standards, ethical AI implementation guidelines, and post-deployment monitoring and evaluation requirements. A balanced approach combining AI's transformative potential with evidence-based strategies, robust governance, and continuous vigilance is necessary to maximize AI's benefits while minimizing potential harms.

Limitations

The study has several limitations, including publication bias, language bias, temporal bias, high statistical heterogeneity, inconsistent outcome measures, variable follow-up periods, context dependence, and quality variability. It also has methodological limitations, such as possible missed relevant studies, selection bias due to multiple reviewers, standardized data extraction with subjectivity in interpretation, and reliance on random-effects models, which assume a normal distribution of effect sizes. The study's comprehensive search strategy may have missed relevant studies, and the inclusion of AI challenges may be highly context-specific.

CONCLUSION

This systematic meta-analysis provides compelling evidence that AI implementation in medical education and healthcare systems is associated with significant challenges that require urgent attention. The high prevalence of algorithmic bias (76%), ethical dilemmas (72%), and implementation barriers (83%) demonstrates that the "dark side" of AI in healthcare is not a theoretical concern but a present reality affecting the majority of AI implementations.

Funding: No funding sources
Conflict of interest: None declared
Ethical approval: Not required

REFERENCES

1. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nat Med. 2022;28(1):31-8.
2. Borkhoff CM, Hawker GA, Kreder HJ, Glazier RH, Mahomed NN, Wright JG. The effect of patients' sex

on physicians' recommendations for total knee arthroplasty. CMAJ. 2024;196(8):E234-E242.
3. Chinta SV, Wang Z, Palikhe A, Zhang X, Kashif A, Smith MA, et al. AI-driven healthcare: Fairness in AI healthcare: A survey. PLOS Digit Health. 2025;4(5):e0000864.
4. Singh H, Meyer AN, Thomas EJ. The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. BMJ Qual Saf. 2024;33(4):330-6.
5. Kaushal A, Altman R, Langlotz C. Geographic distribution of US cohorts used to train deep learning algorithms. JAMA. 2020;324(12):1212-3.
6. Beam AL, Manrai AK, Ghassemi M. Challenges to the reproducibility of machine learning models in health care. JAMA. 2021;325(10):962-3.
7. Chen IY, Pierson E, Rose S, Joshi S, Ferryman K, Ghassemi M. Ethical machine learning in healthcare. Annu Rev Biomed Data Sci. 2023;6:123-44.
8. Lymar L, Kuchyn I, Bielka K, Puljak L. Academic misconduct and artificial intelligence use by medical students, interns and PhD students in Ukraine: a cross-sectional study. BMC Med Educ. 2025;25(1):1496.
9. Ahmed MI, Spooner B, Isherwood J, Lane M, Orrock E, Dennison A. A Systematic Review of the Barriers to the Implementation of Artificial Intelligence in Healthcare. Cureus. 2023;15(10):e46454.
10. Zheng L, Xiao Y. Refining AI perspectives: assessing the impact of ai curricular on medical students' attitudes towards artificial intelligence. BMC Med Educ. 2025;25(1):1115.
11. Wartman SA, Combs CD. Reimagining medical education in the age of AI. Acad Med. 2024;99(3):234-41.
12. Ueda D, Kakinuma T, Fujita S, Kamagata K, Fushimi Y, Ito R, et al. Fairness of artificial intelligence in healthcare: review and recommendations. Jpn J Radiol. 2024;42(1):3-15.
13. Wartman SA, Combs CD. Medical education must move from the information age to the age of artificial intelligence. Acad Med. 2024;99(4):412-9.
14. Tudor Car L, Kyaw BM, Dunleavy G, Smart NA, Semwal M, Rotgans JI, et al. Digital problem-based learning in health professions: systematic review and meta-analysis by the digital health education collaboration. J Med Internet Res. 2022;24(2):e33120.
15. Park YR, Lee Y, Kim J, Kim S, Lee H, Kim JH. The impact of artificial intelligence on diagnostic accuracy in medical education: a longitudinal study. Med Teach. 2024;46(5):623-30.
16. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting algorithmic bias and prediction errors in health care for 1.2 million patients. Science. 2019;366(6464):447-53.
17. Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring Fairness in Machine Learning to Advance Health Equity. Ann Intern Med. 2018;169(12):866-872.
18. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44-56.
19. Krieger K, Hameed I, Quer G, Mack C, Savic M, Mantaj P, et al. Generative pre-trained transformer reinforces historical gender bias in diagnosing women's cardiovascular symptoms. Eur Heart J Digit Health. 2025;ztaf131.
20. Nazer LH, Zatarah R, Waldrip S, Ke JXC, Moukheiber M, Khanna AK, et al. Bias in artificial intelligence algorithms and recommendations for mitigation. PLOS Digit Health. 2023;2(6):e0000278.
21. Cross JL, Choma MA, Onofrey JA. Bias in medical AI: Implications for clinical decision-making. PLOS Digit Health. 2024;3(11):e0000651.
22. Panch T, Mattie H, Celi LA. The "inconvenient truth" about AI in healthcare. NPJ Digit Med. 2019;2:77.
23. Wiens J, Saria S, Sendak M, Ghassemi M, Liang HL, Mellin J, et al. Do no harm: a roadmap for responsible machine learning for health. Nat Med. 2019;25(9):1337-40.
24. Ghassemi M, Naumann T, Schulam P, Beam AL, Chen IY, Ranganath R. A review of challenges and opportunities in machine learning for health. AMIA Jt Summits Transl Sci Proc. 2020;2020:191-200.
25. Char DS, Shah NH, Magnus D. Implementing Machine Learning in Health Care - Addressing Ethical Challenges. N Engl J Med. 2018;378(11):981-3.
26. Larrazabal AJ, Nieto N, Victoria S, Milone DH, Ferrante E. Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis. Proc Natl Acad Sci U S A. 2020;117(23):12592-4.
27. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med. 2019;17(1):195.
28. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence in medicine. Nat Med. 2019;25(1):30-6.
29. Beam AL, Kohane IS. Big Data and Machine Learning in Health Care. JAMA. 2018;319(13):1317-1318.
30. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. Nat Rev Cancer. 2018;18(8):500-14.
31. Masters K. Artificial intelligence in medical education. Med Teach. 2019;41(9):976-982.
32. McCradden MD, Joshi S, Mazwi M, Anderson JA. Ethical limitations of algorithmic fairness in oncology patients. Cancer Med. 2020;9(10):3504-13.
33. Shaw J, Rudzicz F, Jamieson T, Goldfarb A. Artificial Intelligence and the Future of Health Care. Healthc Manage Forum. 2019;32(3):139-44.

34. Challen R, Denny J, Pitt M, Gompels L, Edwards T, Tsaneva-Atanasova K. Artificial intelligence, bias and clinical safety. BMJ Qual Saf. 2019;28(3):231-7.
35. Buolamwini J, Gebru T. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proc Mach Learn Res. 2018;81:1-15.
36. Vayena E, Blasimme A, Tasioulas J. Machine learning in medicine: Addressing ethical challenges. PLoS Med. 2018;15(11):e1002689.
37. Jeong J, Kim S, Pan L, Hwang D, Kim D, Choi J, et al. Reducing the workload of medical diagnosis through artificial intelligence: A narrative review. Medicine (Baltimore). 2025;104(6):e41470.
38. Cabitza F, Rasoini R, Gensini GF. Unintended Consequences of Machine Learning in Medicine. JAMA. 2017;318(6):517-8.
39. Wartman SA, Combs CD. Medical Education must Move from the Information Age to the Age of Artificial Intelligence. Acad Med. 2018;93(8):1107-9.
40. Paranjape K, Schinkel M, Panday RN, Car J, Nanayakkara P. Introducing Artificial Intelligence Training in Medical Education. JMIR Med Educ. 2019;5(2):e16048.
41. Norgeot B, Glicksberg BS, Trupin L, Lituiev D, Williams J, Litchman J, et al. Assessment of a Manifestation of Bias in Real-World Electronic Health Record Data That Impacts Patient Selection and Clinical Trial Design. JAMA Netw Open. 2019;2(10):e1913418.
42. Young AT, Amara D, Bhattacharya A, Wei ML. Patient and general public attitudes towards clinical artificial intelligence: a mixed methods systematic review. Lancet Digit Health. 2021;3(9):e599-e611.
43. Sendak MP, Gao M, Brajer N, Balu S. Presenting machine learning model outcomes to clinicians: Stakeholder-centric design process. JMIR Med Inform. 2020;8(10):e15418.
44. Cai CJ, Winter S, Steiner D, Wilcox L, Terry M. "Hello AI": Uncovering the Onboarding Needs of Medical Practitioners for Human-AI Collaborative Decision-Making. Proc ACM Hum-Comput Interact. 2019;3(CSCW):1-24.
45. Chan S, Teo NWY, Lieow JLM, Ho HY, Somasundaram N, Jeyaraman K, et al. Artificial intelligence in medical education: a review of the literature. Med Teach. 2021;43(12):1351-1358.
46. Mhasawade V, Zhao Y, Chunara R. Machine learning and algorithmic fairness in public and population health. Health Aff (Millwood). 2021;40(10):1586-1594.
47. Rigby MJ. Ethical Dimensions of Using Artificial Intelligence in Health Care. AMA J Ethics. 2019;21(2):E121-124.
48. Vyas DA, Eisenstein LG, Jones DS. Hidden in Plain Sight - Reconsidering the Use of Race Correction in Clinical Algorithms. N Engl J Med. 2020;383(9):874-82.
49. Cirillo D, Catuara-Solarz S, Morey C, Guney E, Subirats L, Mellino S, et al. Sex and gender differences in artificial intelligence framework for health medical and personal data. NPJ Digit Med. 2020;3:81.
50. Chen IY, Szolovits P, Ghassemi M. Can AI Help Reduce Disparities in General Medical Care? AMA J Ethics. 2019;21(2):E167-79.
51. Haleem A, Javaid M, Singh RP, Suman R. Medical 4.0 technologies for healthcare: Features, capabilities, and applications. Internet of Things and Cyber-Physical Systems. 2022;2:12-30.
52. Kuan V, Denaxas S, Gonzalez-Izquierdo A, Direk K, Khorasani-Zadeh O, Fatemifar G, et al. A chronological map of 308 physical and mental health conditions from 21 million people in England during 2001-2018. Lancet Digit Health. 2019;1(2):e63-e77.
53. Lee P, Bubeck S, Petro J. Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine. N Engl J Med. 2023;388(13):1233-1239.
54. McCoy LG, Nagaraj S, Morgado F, Harish V, Das S, Celi LA. What do medical students actually need to know about artificial intelligence? NPJ Digit Med. 2020;3:86.
55. Blease C, Kaptchuk TJ, Bernstein MH, DesRoches CM, Jossel SG. Artificial Intelligence and the Future of Primary Care: Exploratory Qualitative Study of UK General Practitioners' Views. J Med Internet Res. 2019;21(3):e12802.
56. Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng. 2018;2(10):719-731.
57. Sapci AH, Sapci HA. Artificial Intelligence Education and Training Programs for Medical Students and Health Care Professionals: Scoping Review. JMIR Med Educ. 2020;6(2):e19313.
58. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011;343:d5928.
59. Wells GA, Shea B, O'Connell D, Peterson J, Welch V, Losos M, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Ottawa Hospital Research Institute. 2013.
60. Schulz KF, Grimes DA. Sample size calculations in randomised trials: mandatory and mystical. Lancet. 2005;365(9467):1348-53.
61. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ. 2019;366:l4898.
62. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med. 2001;134(8):663-94.
63. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement:

guidelines for reporting observational studies. Lancet. 2007;370(9596):1453-7.
64. Ricci Lara MA, Almansa-Castillo R, Castro C, Pellegrini C, de la Rosa R, Reyes M. Addressing bias in medical machine learning: a systematic review. NPJ Digit Med. 2022;5(1):139.
65. Vollmer S, Mateen BA, Bohner G, Király FJ, Ghani R, Jonsson P, et al. Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. BMJ. 2020;368:l6927.
66. Vayena E, Blasimme A, Cohen IG. Machine learning in medicine: addressing ethical challenges. PLoS Med. 2022;19(4):e1003919.
67. US Food and Drug Administration. Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan. Silver Spring, MD: FDA; 2021. Available at: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices. Accessed on 3 August 2025.
68. European Commission. Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Off J Eur Union. 2024;L1689:1-144.
69. Vollmer S, Mateen BA, Bohner G, Király FJ, Ghani R, Jonsson P, et al. Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. BMJ. 2020;368:l6927.
70. Khan SM, Liu X, Nath S, Korot E, Faes L, Wagner SK, et al. A global review of publicly available datasets for ophthalmological training: the importance of diversity, equity, and inclusion. Lancet Digit Health. 2023;5(6):e396-e407.
71. Celi LA, Cellini J, Charpignon ML, Dee EC, Dernoncourt F, Eberhardt R, et al. Sources of bias in artificial intelligence that perpetuate healthcare disparities. J Med Internet Res. 2022;24(3):e33923.

Cite this article as: Sannathimmappa MB. The dark side of artificial intelligence in medical education and the healthcare system: challenges and strategies for a balanced approach: a systematic meta-analysis. Int J Sci Rep 2026;12(1):41-50.