

Generative AI in Pediatric and Neonatal Intensive Care Units: Tunisian Cross-Sectional Survey

Mariam ben Mbarek^{1,2}, Syrine Bouslama^{1,2}, Seyfeddine Zayani^{1,2}, Abir Daya^{1,2}, Farah Thabet^{1,2},
Slaheddine Chouchane^{1,2}

¹ Department of Pediatrics, Fattouma Bourguiba University Hospital, Monastir, Tunisia. ² University of Monastir, Monastir, Tunisia.

Corresponding author : Seyfeddine Zayani · zayani.seyfeddine@gmail.com · ORCID 0000-0001-7585-1860

ABSTRACT

Background: Generative artificial intelligence (AI) is increasingly used in healthcare, but its use in pediatric and neonatal intensive care remains poorly documented. **Objective:** To assess generative AI use among physicians working in pediatric and neonatal intensive care in Tunisia, focusing on clinical applications, risk-related practices, governance, and training needs. **Methods:** We conducted a national cross-sectional online survey among residents and senior physicians practicing in pediatric intensive care units, neonatal intensive care units, or mixed units in Tunisia. The questionnaire explored AI use, clinical applications, verification practices, patient-data use, perceived safety, institutional rules, and training priorities. A composite AI risk score was calculated among AI users. **Results:** Ninety-four physicians participated, including 79 residents and 15 senior physicians. Overall, 88 (94%) reported medical use of generative AI during the previous 3 months; 72 (82%) used it at least weekly and 83 (94%) reported at least one clinical use. Insufficient verification was reported by 23 users (26%), and 58 (66%) had entered real patient clinical data into an AI tool. AI had influenced a real diagnostic or therapeutic decision in 56/87 respondents (64%). Eighteen users (20%) were classified as high risk. Clear institutional rules were absent for 85 respondents (90%), and only 11 (12%) had received formal training. **Conclusion:** Among surveyed physicians in pediatric and neonatal intensive care in Tunisia, generative AI use was widespread and often clinically influential, contrasting with limited training and governance. Structured education, confidentiality safeguards, and clear local guidance are needed.

Keywords: Artificial Intelligence; Pediatric Intensive Care Units; Physicians; Patient Safety; Survey

INTRODUCTION

Generative artificial intelligence (AI), particularly large language models (LLMs), has rapidly emerged as a potentially transformative technology in health care. These systems are increasingly being explored for a broad range of applications, including knowledge retrieval, documentation support, chart summarization, patient communication, and clinical decision support [1-4].

However, enthusiasm for these tools has been accompanied by substantial concerns regarding their safe integration into clinical practice. Hallucinations, output variability, limited explainability, privacy and confidentiality risks, and unresolved medico-legal and regulatory issues remain important barriers to responsible adoption. Beyond technical performance, several authors have emphasized that successful implementation requires governance, validation, and structured integration strategies adapted to clinical environments [1, 2, 5].

These issues may be particularly relevant in intensive care settings, where cognitive workload is high, unstructured clinical data are abundant, and decisions are often time-sensitive. In such environments, LLMs may appear especially attractive because they could support documentation, case summarization, communication, and elements of clinical

reasoning. At the same time, their use in critical care raises specific concerns related to reliability, inappropriate reliance on generated outputs, and the need for clinician AI literacy [4].

Pediatric-specific data remain scarce. A recent survey among pediatric providers showed that LLMs were already being used in both clinical and nonclinical work, although nonclinical use predominated and important concerns persisted regarding safety, privacy, and institutional guidance [3]. More broadly, early evidence suggests that, in many clinical settings, generative AI still functions primarily as an information-support tool rather than as a truly integrated clinical service technology [2]. Yet little is known about how early-career physicians working in pediatric and neonatal intensive care use these tools in real-world practice, particularly with regard to clinically influential uses, potentially risky practices, institutional governance, and training needs [2, 3].

In Tunisia, university hospitals and specialist centers constitute the tertiary level of care [6]. In parallel, the health system is undergoing progressive digital transformation, with increasing attention to health information systems and digital-health technologies [7]. Personal data processing is governed by a national legal framework and overseen by the National Authority for the Protection of Personal Data [8].

However, the clinical integration of generative AI is still emerging, and specific institutional frameworks governing its use by healthcare professionals remain limited. This evolving context makes it particularly relevant to examine how these tools are already being used in high-acuity pediatric and neonatal settings.

The aim of the present study was therefore to assess the use of generative AI among early-career physicians working in pediatric and neonatal intensive care in Tunisia.

METHODS

Study design and participants

We conducted a national cross-sectional online survey among early-career physicians working in pediatric and neonatal intensive care settings in Tunisia. For the purposes of this study, early-career physicians were defined as residents undergoing specialty training and senior physicians at an early post-specialty academic career stage (assistant hospital-university physicians in the Tunisian academic system). Eligible participants were currently practicing in a pediatric intensive care unit (PICU), neonatal intensive care unit (NICU), or mixed PICU/NICU setting.

Based on estimates obtained from colleagues involved in pediatric and neonatal intensive care across Tunisia, approximately 146 early-career physicians were potentially eligible at the time of the survey, including approximately 108 residents and 38 senior physicians. As no exhaustive national registry of eligible physicians was available, these figures should be considered approximate.

Participation was voluntary and anonymous. Responses were collected over a 15-day period, from March 20 to April 3, 2026. The survey link was disseminated electronically through professional physician networks using WhatsApp and Messenger.

Survey development and dissemination

The questionnaire was developed de novo specifically for this study and was not adapted from a previously validated instrument. Before dissemination, it was reviewed by two professors of pediatrics and pretested in five physicians. The final questionnaire comprised 21 items organized into four main domains: respondent characteristics (5 items), recent use of generative AI (1 item), patterns and practices of generative AI use (12 items), and institutional governance and training needs (3 items). It consisted mainly of closed-ended questions, including single-choice, multiple-choice, and Likert-scale items. The questionnaire was designed to be concise, with an estimated completion time of approximately 4-5 minutes. It was developed and administered using Google Forms and distributed via WhatsApp and Messenger. The complete questionnaire is available from the corresponding author upon request.

Data collection and study variables

The following demographic and professional characteristics were collected: age, sex, professional status (resident or

senior physician), years of experience, and type of unit of practice (PICU, NICU, or mixed PICU/NICU).

Professional seniority was recorded as the number of years spent in the participant's current career stage, corresponding to years of residency for residents and years of assistantship for senior physicians. For senior physicians, previous years of residency were therefore not included in this variable.

Recent use of generative AI was defined as self-reported use of at least one generative AI tool for medical purposes during the previous 3 months. Respondents were also asked about the AI tools they had used, the frequency of use, the main contexts of use, and the clinical tasks for which these tools had been used.

Perceptions regarding AI were explored using Likert-scale items assessing perceived time saving and self-perceived ability to use generative AI safely in medical practice.

Potentially risky practices were assessed through questions addressing verification of AI-generated answers using reliable sources, entry of real patient clinical data into generative AI tools, degree of anonymization of entered data, and reported influence of AI on real diagnostic or therapeutic decisions. Participants were also asked whether they had ever identified an AI-generated response as false, misleading, or potentially dangerous.

At the institutional level, respondents were asked whether clear rules regarding AI use existed in their department or hospital and whether they had received formal training on the medical use of generative AI. Training priorities were explored through multiple-choice items covering confidentiality, critical appraisal, appropriate clinical use, prompting, scientific writing, medico-legal issues, and local governance.

Composite AI risk classification

To provide a simple summary measure of potentially risky AI-related practices, we constructed a study-specific composite AI risk classification among respondents who reported recent use of generative AI. The score was intended as a descriptive tool rather than a validated risk prediction instrument. Four complementary behaviors were selected because they reflected key dimensions relevant to responsible clinical AI use: verification of generated information, use of real patient data, protection of patient confidentiality through anonymization, and influence of AI on real clinical decision-making.

The score ranged from 0 to 4 and included four binary components: insufficient verification of AI-generated answers, defined as reporting that answers were never, rarely, or only sometimes verified using a reliable source; entry of real patient clinical data into a generative AI tool; partial or absent anonymization of entered patient data; and reported influence of AI on a real diagnostic or therapeutic decision. Each component contributed one point. Equal weighting was used because no validated weighting system

exists for these behaviors and to avoid introducing unsupported assumptions regarding their relative importance.

The composite score was calculated only among AI users. For descriptive purposes, participants were classified according to the number of concurrent risk-related practices as low risk (0-1 point), moderate risk (2 points), or high risk (3-4 points). These categories were intended to provide an ordinal descriptive framework and should not be interpreted as validated clinical risk thresholds.

Statistical analysis

Continuous variables were summarized as median and interquartile range (IQR), whereas categorical variables were presented as counts and percentages. The main descriptive analyses were performed in the overall sample and, when appropriate, among respondents who reported recent use of generative AI.

Exploratory comparisons between residents and senior physicians were performed using the Mann-Whitney U test for continuous variables and the chi-square test or Fisher's exact test, as appropriate, for categorical variables. All tests were two-sided, and a p value < 0.05 was considered statistically significant. Missing data were not imputed; percentages were calculated using the number of valid responses for each item when applicable. All analyses were performed using IBM SPSS Statistics, version 26.0.

Ethics

No formal ethics committee review was sought for this study. Participation was voluntary and anonymous, and no identifying personal information or patient-identifiable data were collected.

Use of generative AI in manuscript preparation

During the preparation and revision of this manuscript, the authors used ChatGPT (OpenAI) as an editorial support tool for language refinement, restructuring, and improvement of clarity. Generative AI was not used for data generation or statistical analysis. All scientific content, interpretations, references, and conclusions were critically reviewed and validated by the authors, who take full responsibility for the final content of the manuscript.

RESULTS

A total of 94 physicians completed the survey, representing approximately 64% of the estimated eligible population. Participants included 79 residents (84%) and 15 senior

physicians (16%). The median age was 28 years (IQR 26-30), and the median time in the current career stage was 2 years (IQR 1-4). Most respondents were women (69/94, 73%). Regarding practice setting, 32/94 respondents (34%) worked in a pediatric intensive care unit (PICU), 36/94 (38%) in a neonatal intensive care unit (NICU), and 26/94 (28%) in a mixed PICU/NICU setting. Overall, 88/94 respondents (94%) reported having used generative AI for medical purposes during the previous 3 months. Formal training in the medical use of generative AI had been received by only 11/94 respondents (12%), and 85/94 (90%) reported the absence of clear institutional rules regarding AI use (Table I).

Among AI users, 83/88 (94%) reported at least one clinical application. The most frequently reported clinical uses were explaining pathophysiological mechanisms (52/88, 59%), generating a differential diagnosis (46/88, 52%), and discussing treatment or dosage (44/88, 50%). Eighty users (91%) agreed that generative AI saved time in their work, whereas 46/88 (52%) agreed that they felt able to use it safely in medical practice (Table II).

Among AI users, 23 (26%) reported insufficient verification of AI-generated answers, and 58 (66%) had entered real patient clinical data into a generative AI tool. Among those entering patient data, 8/58 (14%) reported partial or absent anonymization. AI had influenced a real diagnostic or therapeutic decision in 56/87 respondents (64%), and 54/88 (61%) reported having identified at least one AI-generated response as false, misleading, or potentially dangerous. Based on the composite classification, 18/88 users (20%) were classified as high risk (Table III).

In an exploratory comparison among AI users, incomplete anonymization of entered clinical data was more frequently reported by senior physicians than by residents (27% vs 6%, Fisher's exact $p = 0.026$).

Exploratory comparisons of generative AI use according to professional status are presented in Supplementary Table S1. Overall patterns were broadly similar between residents and senior physicians. Senior physicians more frequently reported using generative AI for presentations or teaching (60% vs 29%, Fisher's exact $p = 0.034$) and for preparing explanations for parents or ward rounds (33% vs 8%, Fisher's exact $p = 0.019$). Given the small number of senior physicians and the multiple exploratory comparisons performed, these findings should be interpreted cautiously (Supplementary Table S1).

Table I. Baseline characteristics of the respondents

Variable	Total (N=94)	Residents (n=79)	Senior physicians (n=15)	p value
Age, years, median (IQR)	28 (26-30)	28 (26-29)	33 (31-34)	<0.001
Years in current career stage, median (IQR)	2 (1-4)	2 (1-4)	2 (1-3)	0.789
Female sex, n (%)	69 (73)	59 (75)	10 (67)	0.534
PICU, n (%)	32 (34)	27 (34)	5 (33)	0.850
NICU, n (%)	36 (38)	31 (39)	5 (33)	
Mixed PICU/NICU, n (%)	26 (28)	21 (27)	5 (33)	
Use of generative AI during the last 3 months, n (%)	88 (94)	73 (92)	15 (100)	0.585
Formal training in medical use of generative AI, n (%)	11 (12)	8 (10)	3 (20)	0.374
Absence of clear institutional rules regarding AI use, n (%)	85 (90)	71 (90)	14 (93)	0.546

IQR, interquartile range; PICU, pediatric intensive care unit; NICU, neonatal intensive care unit.

Table II. Modalities of generative AI use among users (n = 88)

Variable	n (%)
Tools used	
ChatGPT	85 (97%)
Gemini	36 (41%)
Perplexity	11 (12%)
Claude	8 (9%)
Copilot	2 (2%)
Other	13 (15%)
Frequency of use	
About once a month	1 (1%)
Two to three times a month	15 (17%)
One to two times a week	17 (19%)
Several times a week	30 (34%)
Daily	25 (28%)
At least weekly use	72 (82%)
Main contexts of use	
Bibliographic search / guidelines	70 (79%)
Translation / rephrasing	44 (50%)
Scientific writing	42 (48%)
Clinical reasoning support	38 (43%)
Presentations / teaching	30 (34%)
Medical documentation / drafting	20 (23%)
Other context	2 (2%)
Clinical tasks	
Explaining pathophysiological mechanisms	52 (59%)
Generating a differential diagnosis	46 (52%)
Discussing treatment or dosage	44 (50%)
Suggesting a therapeutic approach	28 (32%)
Rephrasing a management plan	19 (22%)
Summarizing a clinical case	17 (19%)
Interpreting laboratory tests / blood gases / imaging	11 (12%)
Preparing explanations for parents or ward rounds	11 (12%)
No clinical use	5 (6%)
At least one clinical use	83 (94%)
Perceptions	
Agreed or strongly agreed that AI saves time	80 (91%)
Agreed or strongly agreed that they felt able to use AI safely	46 (52%)

Multiple responses were allowed for tools, contexts of use, and clinical tasks.

Table III. Risk-related practices, governance, and training priorities regarding generative AI

Variable	n (%)
Verification of AI-generated responses among AI users (n = 88)	
Never verified AI-generated answers	1 (1%)
Rarely verified AI-generated answers	2 (2%)
Sometimes verified AI-generated answers	20 (23%)
Often verified AI-generated answers	34 (39%)
Always verified AI-generated answers	31 (35%)
<i>Insufficient verification</i>	23 (26%)
Confidentiality and anonymization among AI users (n = 88)	
Entered real patient clinical data into a generative AI tool	58 (66%)
Partial or absent anonymization among all AI users	8 (9%)
Partial or absent anonymization among those who entered real patient data	8/58 (14%)
Clinical decision influence and identification of erroneous outputs	
AI influenced a real diagnostic or therapeutic decision	56/87 (64%)
Identified at least one AI-generated response as false, misleading, or dangerous	54/88 (61%)
Did not know whether such an error had occurred	10/88 (11%)
Composite AI risk classification among AI users (n = 88)	
Score 0	15 (17%)
Score 1	21 (24%)
Score 2	34 (39%)
Score 3	16 (18%)
Score 4	2 (2%)
<i>Low risk</i>	36 (41%)
<i>Moderate risk</i>	34 (39%)
<i>High risk</i>	18 (20%)
Institutional governance and formal training in the overall sample (N = 94)	
No clear institutional rules regarding AI use	85 (90%)
Clear institutional rules regarding AI use	4 (4%)
Did not know whether rules existed	5 (5%)
Formal training in medical use of generative AI	11 (12%)
Training priorities in the overall sample (N = 94)	
Verification of answers and critical appraisal	55 (59%)
Appropriate clinical use / situations where AI should not be used	50 (53%)
Confidentiality and data protection	46 (49%)
Scientific and academic writing	46 (49%)
Medico-legal and ethical issues	37 (39%)
Effective prompting for medical use	34 (36%)
Limitations and hallucinations of AI models	17 (18%)
Development of a local or institutional charter	17 (18%)
No training needed	5 (5%)

Percentages are calculated among AI users (n = 88) or the overall sample (N = 94), as indicated. For AI influence on a real diagnostic or therapeutic decision, one response was missing (n = 87). The composite AI risk score ranged from 0 to 4.

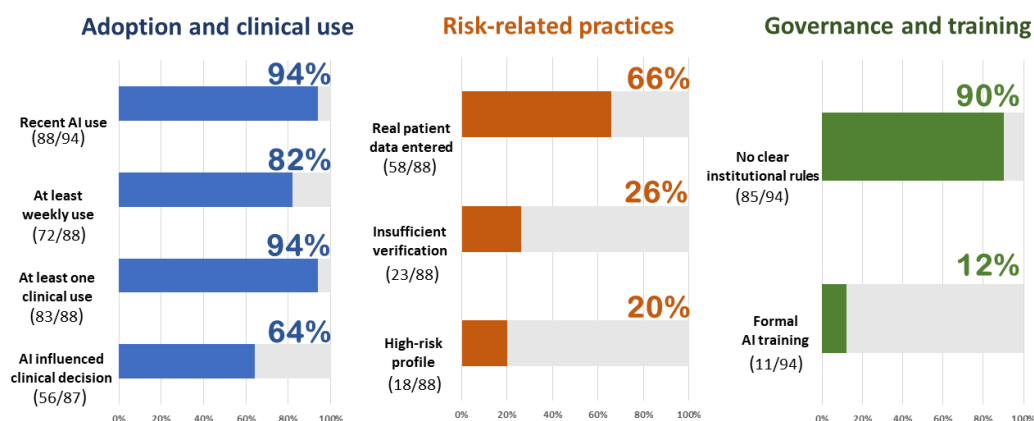


Figure 1. Adoption-risk-governance profile of generative AI use among early-career physicians working in pediatric and neonatal intensive care in Tunisia. Indicators are grouped into three domains: adoption and clinical use, risk-related practices, and governance/training. Percentages are reported for the overall sample (N = 94) or among AI users (n = 88), as applicable; AI influence on a real diagnostic or therapeutic decision was based on 87 valid responses.

DISCUSSION

In this national cross-sectional survey, we found that generative AI use was already widespread among early-career physicians working in pediatric and neonatal intensive care. Use was not only frequent, but also often clinically embedded. At the same time, this rapid uptake coexisted with limited formal training, minimal institutional governance, and several potentially risky practices. Taken together, these findings suggest that, in this population, generative AI is no longer a marginal or purely academic tool, but a technology already integrated into everyday practice without a matching level of structured oversight or preparation [1, 4, 5, 9, 10].

Our findings extend previous survey-based literature. In the pediatric provider survey by Kisvarday et al., roughly half of respondents had used an LLM, and nonclinical use predominated over clinical use [3]. In contrast, in our study, 94% of respondents reported recent use and 94% of AI users reported at least one clinical application. Although differences in setting, population, and timing preclude direct comparison, this contrast suggests that uptake may be particularly rapid among early-career physicians in high-workload environments such as PICU and NICU, where fast access to synthesized information, text generation, and cognitive support may be perceived as especially valuable [3, 4, 9, 10].

A key finding is that AI use was not confined to bibliographic search, translation, or writing support. Many respondents reported using these tools for differential diagnosis, treatment or dosage discussion, and therapeutic suggestion, and nearly two thirds stated that AI had already influenced a real diagnostic or therapeutic decision. This is important because early reviews have suggested that current clinical

use of generative AI remains more informational than truly integrated into care processes [2, 9, 10]. Our results suggest that, at least from the perspective of early-career physicians, these tools are already shaping clinical reasoning in real-world practice. This shift from informational assistance toward clinically influential use deserves careful attention.

This point is particularly relevant in light of automation bias. Khera et al. emphasized the risk that assistive AI may influence clinician judgment in ways that are not always recognized or adequately checked [11]. In our study, 64% of respondents reported that AI had influenced a real decision, while 61% had already encountered an AI-generated response they considered false, misleading, or potentially dangerous. This coexistence is noteworthy. It suggests that awareness of model fallibility does not necessarily prevent clinically influential use. Rather, early-career physicians may already be using these tools as cognitive aids before clear norms for calibration of trust, verification, and appropriate clinical boundaries have been established [11, 12].

Another important finding was the gap between perceived utility and perceived safety. Although most users considered generative AI time-saving, only about half felt able to use it safely in medical practice. This suggests that perceived practical benefit may be developing faster than competence in critical appraisal and safe supervision of AI-generated outputs. As highlighted by McCoy et al., clinicians increasingly need skills in evaluating, contextualizing, and verifying AI-generated information rather than simply accessing it [12].

Confidentiality and governance also emerged as important concerns. Two thirds of AI users reported entering real patient clinical data into generative AI tools, while 14% of those entering such data reported partial or absent anonymization. These practices occurred in a setting where

90% of respondents reported the absence of clear institutional rules regarding AI use. This finding reinforces concerns raised in the literature regarding the need for privacy safeguards, institutional oversight, and explicit governance frameworks for clinical generative AI use [1, 5, 13].

These findings are also relevant in the broader context of low- and middle-income health systems. Generative AI may offer particular opportunities in resource-constrained settings by facilitating access to medical information and supporting healthcare professionals facing workforce and resource limitations [14]. At the same time, important barriers to responsible AI implementation in such settings have been highlighted, including limited digital infrastructure, scarcity of locally representative data, regulatory gaps, privacy and data-security concerns, and insufficient local expertise and training [14, 15]. In this context, the combination observed in our study – rapid adoption, clinically influential use, and limited institutional governance – underscores the importance of developing locally adapted guidance, governance frameworks, and capacity-building strategies rather than relying solely on models developed in high-income settings [15, 16].

The composite AI risk score should be interpreted as a pragmatic descriptive tool rather than a validated instrument. Nevertheless, it helped illustrate an important pattern: nearly 60% of AI users fell into the moderate- or high-risk categories, and one in five was classified as high risk. In our view, the value of this score lies less in precise quantification than in its ability to summarize the coexistence of four clinically relevant behaviors: insufficient verification, entry of real patient data, incomplete anonymization, and AI influence on real decisions. This distinction matters because high adoption alone is not necessarily problematic; high adoption combined with weak governance, limited training, and clinically influential use is what creates the more important challenge.

Our findings also have educational implications. The most frequently selected training priorities were critical appraisal of AI-generated answers, appropriate clinical use, confidentiality, and scientific writing. This suggests that respondents primarily sought support for safe and critical use rather than greater technical sophistication. This is consistent with McCoy et al., who emphasized the need for clinician competencies in hallucinations, prompt sensitivity, bias, uncertainty, and verification [12]. In parallel, Ba et al. showed that ChatGPT-assisted teaching may improve selected practical skills in pediatric trainees when used as a complement to conventional teaching [17]. Together, these findings support integrating generative AI education into

structured training focused on critical appraisal, confidentiality, and appropriate clinical boundaries.

This study has limitations. Its cross-sectional design does not allow causal inference or assessment of temporal trends. Data were self-reported and may therefore be affected by recall bias and social desirability bias. Recruitment through professional social media networks may have favored more digitally engaged respondents and may have overestimated adoption. In addition, the relatively small number of senior physicians limited the robustness of subgroup comparisons. The modest overall sample size, particularly the limited number of participants classified as high risk, also precluded robust multivariable analyses of factors independently associated with risk-related AI practices. Finally, the composite AI risk score was study-specific and should be interpreted as a descriptive summary rather than a validated metric.

Despite these limitations, this study provides one of the first national snapshots of real-world generative AI use among early-career physicians in pediatric and neonatal intensive care. It shows not only that these tools are widely used, but also that they are already influencing clinically relevant thinking in a context where governance and training remain limited. The priority now is therefore not simply to acknowledge the emergence of generative AI, but to organize its safe use through local guidance, confidentiality safeguards, and structured education in verification and critical appraisal.

CONCLUSION

In this national survey of early-career physicians working in pediatric and neonatal intensive care in Tunisia, generative AI use was widespread and extended beyond academic support into clinically influential domains. Within this population, adoption appeared to be occurring faster than the development of formal training and institutional governance. These findings highlight the need for locally adapted guidance, confidentiality safeguards, and structured training in verification and critical appraisal to support the safe and responsible use of generative AI in pediatric and neonatal critical care.

Authors' contributions. SZ, AD initiated the study. SC and FT supervised the study and participated in the interpretation of results and manuscript preparation. MBM and SB conducted statistical analysis and contributed to result interpretation. All authors read and approved the final manuscript. **Funding.** The authors declare that no funds, grants, or other support were received during the preparation of this manuscript. **Conflicts of interest.** The authors have no relevant financial or non-financial interests to disclose. **Co-authors.** Mariem ben Mbarek (Maryemb85@gmail.com); Syrine Bouslama (bouslamasyrine9@gmail.com); Abir Daya (Daya_abir@hotmail.fr); Farah Thabet (ORCID 0000-0003-4824-5691); Slaheddine Chouchane (chouchaneslaheddine5000@gmail.com).

REFERENCES

1. Reddy S. Generative AI in healthcare: an implementation science informed translational path on application, integration and governance. *Implement Sci.* 2024;19(1):27.
2. Yim D, Khuntia J, Parameswaran V, Meyers A. Preliminary evidence of the use of generative AI in health care clinical services: systematic narrative review. *JMIR Med Inform.* 2024;12:e52073.
3. Kisvarday S, Yan A, Yarahuan J, Kats DJ, Ray M, Kim E, et al. ChatGPT use among pediatric health care providers: cross-sectional survey study. *JMIR Form Res.* 2024;8:e56797.
4. Biesheuvel LA, Workum JD, Reuland M, van Genderen ME, Thorat P, Dongelmans D, et al. Large language models in critical care. *J Intensive Med.* 2025;5(2):113-118.
5. Freyer O, Wiest IC, Kather JN, Gilbert S. A future role for health applications of large language models depends on regulators enforcing safety standards. *Lancet Digit Health.* 2024;6(9):e662-e672.
6. United Nations Children's Fund. Tunisia country brief: community health: policy and implementation landscape mapping in the Middle East and North Africa region 2024. Amman: UNICEF Middle East and North Africa Regional Office; 2024.
7. World Bank. Tunisia Health System Strengthening Project (P507209). Washington (DC): World Bank; 2025.
8. Instance Nationale de Protection des Données Personnelles. Loi organique n° 2004-63 du 27 juillet 2004 portant sur la protection des données à caractère personnel. Tunis: INPDP; 2004.
9. Yitzhaki S, Peled N, Kaplan E, Kadmon G, Nahum E, Gendler Y, et al. Comparing ChatGPT-4 and a paediatric intensive care specialist in responding to medical education questions: a multicenter evaluation. *J Paediatr Child Health.* 2025;61(7):1084-1089.
10. Lucas HC, Upperman JS, Robinson JR. A systematic review of large language models and their implications in medical education. *Med Educ.* 2024;58(11):1276-1285.
11. Khera R, Simon MA, Ross JS. Automation bias and assistive AI: risk of harm from AI-driven clinical decision support. *JAMA.* 2023;330(23):2255-2257.
12. McCoy LG, Ng FYC, Sauer CM, Yap Legaspi KE, Jain B, Gallifant J, et al. Understanding and training for the impact of large language models and artificial intelligence in healthcare practice: a narrative review. *BMC Med Educ.* 2024;24(1):1096.
13. Temsah MH, Aljamaan F, Malki KH, Alhasan K, Altamimi I, Aljarbou R, et al. ChatGPT and the future of digital health: a study on healthcare workers' perceptions and expectations. *Healthcare (Basel).* 2023;11(13):1812.
14. Ong JCL, Ning Y, Yang R, Bitterman DS, Liu X, Tham YC, et al. Large language models in global health. *Nature Health.* 2026;1:35-47.
15. World Health Organization. Artificial intelligence for health: supporting countries to deploy responsible AI technologies to accelerate equitable health for all. Geneva: World Health Organization; 2024.
16. Muralidharan V, Ng MY, AlSalamah S, Pujari S, Kalra K, Singh R, et al. Global Initiative on AI for Health (GI-AI4H): strategic priorities advancing governance across the United Nations. *npj Digit Med.* 2025;8(1):219.
17. Ba H, Zhang L, Yi Z. Enhancing clinical skills in pediatric trainees: a comparative study of ChatGPT-assisted and traditional teaching methods. *BMC Med Educ.* 2024;24(1):558.

Supplementary Table S1. Exploratory comparison of generative AI use among residents (n = 73) and senior physicians (n = 15)

Variable	Residents (n = 73)	Senior physicians (n = 15)	p value
Tools used			
ChatGPT	70/73 (96%)	15/15 (100%)	1.000
Gemini	29/73 (40%)	7/15 (47%)	0.774
Perplexity	7/73 (10%)	4/15 (27%)	0.088
Claude	6/73 (8%)	2/15 (13%)	0.620
Copilot	1/73 (1%)	1/15 (7%)	0.313
Other	11/73 (15%)	2/15 (13%)	1.000
Frequency of use			
At least weekly use	61/73 (84%)	11/15 (73%)	0.461
Daily use	20/73 (27%)	5/15 (33%)	0.755
Main contexts of use			
Bibliographic search / guidelines	58/73 (79%)	12/15 (80%)	1.000
Translation / rephrasing	36/73 (49%)	8/15 (53%)	1.000
Scientific writing	33/73 (45%)	9/15 (60%)	0.397
Clinical reasoning support	31/73 (42%)	7/15 (47%)	0.782
Presentations / teaching	21/73 (29%)	9/15 (60%)	0.034
Medical documentation / drafting	17/73 (23%)	3/15 (20%)	1.000
Other context	1/73 (1%)	1/15 (7%)	0.313
Clinical tasks			
Explaining pathophysiological mechanisms	43/73 (59%)	9/15 (60%)	1.000
Generating a differential diagnosis	38/73 (52%)	8/15 (53%)	1.000
Discussing treatment or dosage	34/73 (47%)	10/15 (67%)	0.257
Suggesting a therapeutic approach	22/73 (30%)	6/15 (40%)	0.545
Rephrasing a management plan	17/73 (23%)	2/15 (13%)	0.508
Summarizing a clinical case	14/73 (19%)	3/15 (20%)	1.000
Interpreting laboratory tests / blood gases / imaging	10/73 (14%)	1/15 (7%)	0.682
Preparing explanations for parents or ward rounds	6/73 (8%)	5/15 (33%)	0.019
No clinical use	4/73 (5%)	1/15 (7%)	1.000
At least one clinical use	69/73 (95%)	14/15 (93%)	1.000
Perceptions			
Agreed or strongly agreed that AI saves time	65/73 (89%)	15/15 (100%)	0.342
Agreed or strongly agreed that they felt able to use AI safely	35/73 (48%)	11/15 (73%)	0.092

Data are n/N (%). Comparisons were performed using two-sided Fisher's exact tests. Analyses were exploratory and p values were not adjusted for multiple comparisons. Multiple responses were allowed for tools, contexts of use, and clinical tasks.