Cureus
Part of SPRINGER NATURE

Open Access Original Article

Widespread Use, Minimal Training: Generative Artificial Intelligence Among Undergraduate and Postgraduate Medical Students

Asra Malik [1], Muzaffer Pukhta [1]

1. Pharmacology and Therapeutics, Government Medical College, Srinagar, IND

Corresponding author: Asra Malik, asramalik0@gmail.com

Review began 11/08/2026
Review ended 11/18/2026
Published 11/25/2026

© Copyright 2026
Malik et al. This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

DOI: 10.7759/cureus.120278

Abstract

Introduction: Generative artificial intelligence (AI) has entered medical education faster than curricula or institutional policy have been able to respond. We set out to quantify how far student use of these tools has outpaced formal instruction in them and to test whether engagement differs between undergraduate (UG) and postgraduate (PG) trainees.

Materials and methods: We conducted a cross-sectional, questionnaire-based study of medical students at a tertiary-care medical college in North India between March and May 2025. A structured, self-administered instrument covering knowledge, attitude, and practice was distributed electronically. Groups were compared using the Pearson chi-square or Fisher exact test and independent-samples t-tests, with Benjamini-Hochberg control of the false-discovery rate.

Results: Of 205 responses, 202 consenting respondents were analyzed (116 UG, 86 PG; mean age 25.8 years). Awareness of AI in medicine was near-universal (191/201, 95.0%), and 173/202 (85.6%) had used these tools for study, yet only 46/202 (22.8%) had ever received formal training in them; 92/202 (45.5%) already used AI to answer clinical questions. Attitudes were positive (165/202, 81.7% agreed AI is useful) but tempered by ethical concern (138/201, 68.7%). No statistically robust difference was detected between UG and PG students on any indicator after correction for multiple testing, although the confidence intervals are wide and do not exclude differences of up to 15-25 percentage points.

Conclusions: Substantial self-directed use coexists with a training gap that leaves more than three-quarters of students without any formal instruction, and we detected no robust difference between training levels, although the study was not powered to establish equivalence. The priority is AI literacy and ethics teaching across the medical education continuum rather than provision targeted at one cohort.

Categories: Medical Education, Medical Simulation, Healthcare Technology
Keywords: ai literacy, artificial intelligence, chat gpt, cross-sectional studies, curriculum development, india, knowledge attitude practice, large language models (llm), medical education, medical students

Introduction

The rapid diffusion of generative artificial intelligence (AI), in particular large language models (LLMs) such as ChatGPT, has changed how medical students acquire, process, and apply knowledge [1]. Unlike earlier digital resources, these systems provide interactive, on-demand assistance that students now use to summarize dense material, generate study aids, answer multiple-choice questions, and rehearse clinical reasoning [2,3]. Their performance on standardized medical examinations has accelerated uptake, as shown by an LLM that reached or approached the passing threshold of the United States Medical Licensing Examination without specialized training [2]. AI is simultaneously expanding across clinical medicine, from diagnostic imaging to decision support, raising the stakes of how the next generation of physicians learn to use it [4].

The central problem is not adoption but the absence of accompanying instruction. The opportunities come with well-documented risks, including factual inaccuracy ("hallucination"), opaque reasoning, bias, and the potential erosion of independent critical thinking. These risks make unguided use a concern in a discipline where errors carry clinical consequences [1,5]. AI-generated medical text has also been shown to contain high rates of fabricated and inaccurate citations [6]. These are precisely the failure modes that formal instruction is designed to guard against. When heavy use is not matched by training, students acquire the tool without the competencies needed to audit it, and the resulting gap between what learners do and what they have been taught to do is the actionable target for curriculum design and institutional policy [7,8]. The knowledge, attitude, and practice (KAP) framework offers a structured way to measure that gap: knowledge captures awareness, conceptual understanding, and prior training; attitude captures perceived usefulness, trust, and ethical concern; and practice captures actual usage patterns [9,10].

How such provision should be targeted depends on whether the gap is uniform or concentrated in one cohort. A common assumption in curriculum debates is that AI engagement should track the stage of training. Undergraduate (UG) students, whose primary task is acquiring foundational knowledge and passing examinations, are expected to adopt AI broadly for efficiency, whereas postgraduate (PG) residents, who carry direct patient care responsibility, are expected to be more cautious and selective because the consequences of an erroneous AI output are greater [11,12]. If correct, this would justify stage-specific educational strategies. A recent survey of medical trainees in the United States reported that patterns of generative AI use did vary with the phase of training, lending some support to this view [13]. Before examining the data, we stated a directional expectation: that, relative to UGs, PGs would use AI less, express greater ethical concern, and be less willing to recommend it. This expectation was formulated a priori but was not recorded in a protocol or prospectively registered, and the study is therefore reported as exploratory. Direct UG-PG comparisons within a single institution using one instrument remain uncommon, and the two Indian studies that have made them reached different conclusions: one found PGs clearly ahead in AI readiness, while a more recent tertiary-center survey found broadly similar KAPs [14,15]. Much of the remaining literature pools heterogeneous learners or studies a single level [9,16]. Evidence specific to the Indian context, though now emerging [17-20], remains limited, particularly from Jammu and Kashmir, as many earlier surveys predate the public release of ChatGPT in late 2022 and may not reflect current behavior [21].

We therefore set out to quantify the distance between generative AI use and formal training among medical students at a single tertiary care institution and to establish whether that distance differs by stage of training, framing the UG-PG contrast as a hypothesis to be tested rather than a foregone conclusion.

Materials And Methods

This study is reported in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidance for cross-sectional studies [22] and the Checklist for Reporting Results of Internet E-Surveys (CHERRIES) [23].

Study design and setting

We performed a single-center, cross-sectional, questionnaire-based study at Government Medical College Srinagar, Jammu and Kashmir, India, a tertiary care teaching institution providing undergraduate (UG; Bachelor of Medicine, Bachelor of Surgery (MBBS)) and postgraduate (PG; MD/MS/DNB) medical training. Data were collected over three months in 2025 (March to May).

Participants and sampling

Eligible participants were medical students enrolled at Government Medical College Srinagar during the study period, comprising two groups defined in advance: (i) UG students (MBBS students across academic years and rotating interns) and (ii) PG students enrolled in a postgraduate degree program. For context on coverage, the sanctioned intake at the time was 180 MBBS places per academic year and 259 PG places, so the analytic sample represents a small fraction of those eligible. The instrument asked respondents to name their institution; three of the 202 respondents named another institution (one each from a second medical college in the territory, a medical college in Jammu, and a college outside India) and were retained in the primary analysis, with a sensitivity analysis excluding them reported below. A nonprobability convenience sample was recruited through institutional digital communication channels and messaging applications, through which the questionnaire link was distributed to eligible students. Participation was voluntary and anonymous.

Instrument

The self-administered instrument was developed for this study and organized into four sections: (A) sociodemographic information (age, gender, level of study, and institution); (B) knowledge (awareness of AI in medicine, familiarity with specific tools, perceived applications, and prior formal training); (C) attitude (perceived usefulness, views on whether AI can replace traditional resources, ethical concerns, and support for curricular integration); and (D) practice (prior use, frequency of use, purposes of use, and willingness to recommend AI to peers). Items used dichotomous, three-option, five-point Likert, and multiple-response ("tick all that apply") formats. A pilot study with five third- and fourth-year medical students, who were excluded from the final sample, assessed clarity and usability and led to refinement and shortening of the questionnaire. The full instrument is reproduced in the appendices.

Composite KAP scoring

To summarize each domain, we constructed transparent composite scores, with each component scoring one point. The knowledge score (0-4) summed awareness of AI in medicine, prior formal training, familiarity with three or more named tools, and selection of at least one of the four listed applications rather than "I don't know"; this last component therefore distinguishes respondents who could name any application from those who could name none, and no judgment is made about whether a given use is pedagogically desirable. One of the listed tools, Google DeepMind, is a research organization rather than a tool students can use; 11

respondents named it, and removing it from the tool count changes the three-or-more-tools component for no respondent. The attitude score (0-3) summed agreement that AI is useful, support for curricular integration, and willingness to recommend AI. Awareness of ethical issues was deliberately excluded from this composite because it is a caution rather than an endorsement, so adding it would make a higher score ambiguous; it is reported as a separate item. The practice score (0-4) summed ever-use, regular (daily or frequent) use, use for one or more purposes, and use for two or more purposes. Each domain was expressed as a percentage of its maximum. Internal consistency (Kuder-Richardson 20) was 0.43 for knowledge, 0.71 for the three-item attitude score, and 0.70 for practice; the knowledge composite was the weakest, and all three were treated as exploratory descriptive summaries rather than validated scales.

Data collection

The questionnaire was administered electronically using Google Forms (Google LLC, Mountain View, CA, USA). The landing page presented the study aims and an informed consent statement; respondents indicating agreement proceeded to the items. No personal identifiers were collected, and responses were recorded automatically in a restricted-access spreadsheet available only to the investigators. The instrument comprised 18 items on a single adaptive form; only the consent item was mandatory, duplicate submissions were limited to one response per Google account, and no IP addresses or cookies were used. Because the link was forwarded through messaging channels, a participation rate could not be determined.

Statistical analysis

Data were analyzed in Python 3.13 (Python Software Foundation, Wilmington, DE, USA) using NumPy 2.5, pandas 3.1, and SciPy 1.18; regression models were fitted in statsmodels 0.15. Categorical variables are summarized as counts, percentages, and 95% confidence intervals (CIs; Wilson method); age is summarized as mean ± standard deviation (SD), median, and range. UG and PG groups were compared on the affirmative-category proportion using the Pearson chi-square test with the Yates continuity correction, with the Fisher exact test used when any expected cell count was below five. Effect sizes (phi coefficient or Cramér V) are computed from the uncorrected statistic, so they are not algebraically consistent with the test statistics. Between-group differences in proportions are reported in percentage points with 95% Wald CIs. The ordered frequency-of-use item was analyzed using the Mann-Whitney U test and, like every other frequency analysis, was restricted to respondents who reported ever using AI. Continuous variables were compared using independent-samples t-tests (Welch correction) and Cohen's d. Because multiple comparisons were performed, P values for the UG-PG family (nine categorical items plus three composites) were examined under Benjamini-Hochberg control of the false discovery rate, and adjusted q values are reported. Because training level and age were strongly collinear (point-biserial $r = 0.58$), the two nominal level signals were reexamined in multivariable models (binary logistic and linear regression) including training level, age, and gender; each model used the 201 respondents with complete age data. A further model of ever-use on age was fitted after inspecting the data and is reported as post hoc. A two-sided $P < 0.05$ was considered nominally significant. Given the achieved sample (116 UG, 86 PG), the study had approximately 80% power to detect a between-group difference of roughly 18-20 percentage points in a dichotomous item or a standardized mean difference of $d \approx 0.40$; smaller differences could not be reliably detected, and no equivalence margin was set in advance, so the study cannot demonstrate equivalence.

Ethics

The study was reviewed by the Institutional Review Board of Government Medical College Srinagar and granted exemption from full ethical review in accordance with institutional guidelines (reference no. IRBGMC-SGR/Pharma/1492, dated February 13, 2025) before data collection began because it was an anonymous, voluntary, self-administered educational survey that collected no personal identifiers and posed no foreseeable risk to participants. Electronic informed consent was obtained from every respondent before any item was presented, and data were used solely for research, in accordance with the Declaration of Helsinki [24]. Age was self-reported in a free-text field. One first-year respondent entered an age of 16, which is below the minimum age for admission to the MBBS program and is most likely a typing error; because the survey was anonymous and collected no identifiers, the entry could not be verified or corrected. No other respondent reported an age below 18. That record is retained in the primary analysis and excluded in a sensitivity analysis reported below.

Results

Participant characteristics

Of 205 questionnaire responses, three respondents declined consent, leaving 202 consenting respondents in the analytic sample (Figure *1*). Of these, 116 (57.4%) were UGs, and 86 (42.6%) were PG residents. The mean age was 25.8 ± 4.5 years (median 26, range 16-42). As expected from the two training stages, UGs were substantially younger than PGs (23.6 ± 4.0 vs 28.8 ± 3.2 years; $t = -10.4$, $P < 0.001$, Cohen's $d = -1.43$) (Figure *2*). Respondents were predominantly female (120/201, 59.7%). Full characteristics are shown in Table *1*.

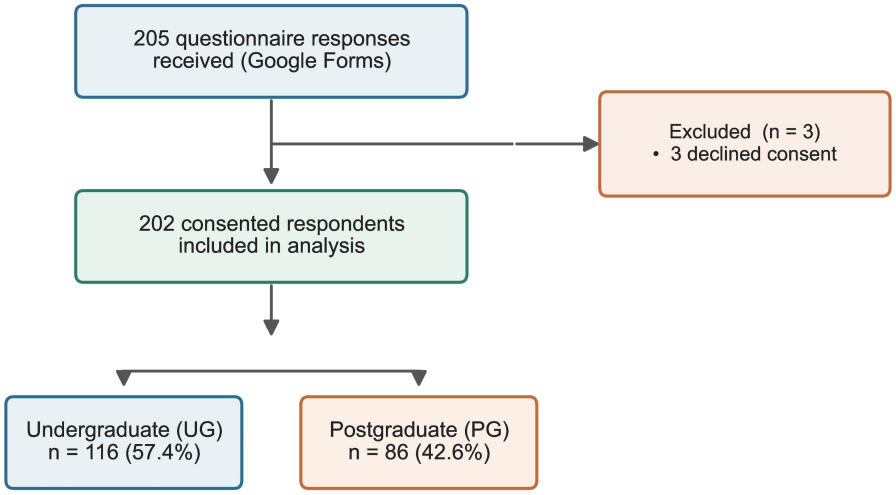


FIGURE 1: Participant recruitment and the analytic sample

Flow of responses from receipt to the analytic sample. Of 205 responses received, three were excluded because consent was declined, leaving 202 respondents: 116 UG and 86 PG.

UG, undergraduate; PG, postgraduate

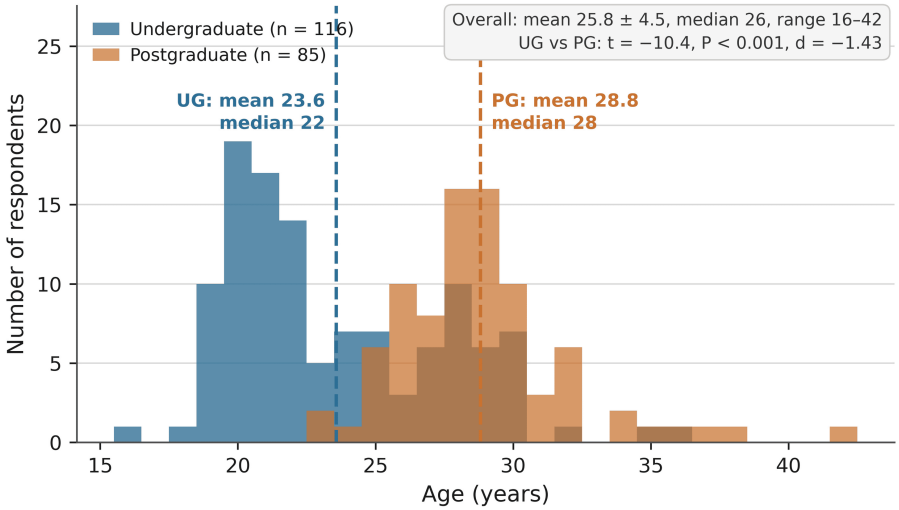


FIGURE 2: Age distribution by training level

UGs (including interns) cluster at younger ages (mean 23.6, median 22 years) and PG residents at older ages (mean 28.8, median 28 years). Dashed lines mark group means; medians are also shown.

UG, undergraduate; PG, postgraduate

Variable	Category	n / N	Value or % (95% CI)
Age, years	Mean ± SD (overall)	201	25.8 ± 4.5
	UG	116	23.6 ± 4.0
	PG	85	28.8 ± 3.2
Gender	Female	120/201	59.7 (52.8-66.2)
	Male	80/201	39.8 (33.3-46.7)
	Prefer not to say	1/201	0.5 (0.1-2.8)
Level of study	UG	116/202	57.4
	PG	86/202	42.6
Year of study (UG)	First year	40/116	34.5
	Second year	44/116	37.9
	Third year	9/116	7.8
	Final year	9/116	7.8
	Intern	14/116	12.1

TABLE 1: Demographic characteristics of the analytic sample (n = 202)

One PG did not report age (age denominators: 85 for PG, 201 overall). Year-of-study percentages use the 116 UGs as the denominator; 84 of 116 (72.4%) UGs were in the first or second year.

SD, standard deviation; UG, undergraduate; PG, postgraduate; CI, confidence interval

Knowledge and awareness

Awareness of AI in medicine was near-universal: 191/201 (95.0%, 95% CI 91.1-97.3) had heard of AI in healthcare or medical education. Familiarity was concentrated on a small number of consumer LLMs, with ChatGPT (194/202, 96.0%) dominant; domain-specific clinical tools such as UpToDate AI were rarely named (5/202, 2.5%) (Figure *3*). Most respondents recognized legitimate applications of AI, including learning and summarizing topics (180/202, 89.1%), writing assignments (155/202, 76.7%), and diagnostic assistance (130/202, 64.4%). Despite this high awareness, formal training was scarce: only 46/202 (22.8%, 95% CI 17.5-29.0) had attended any workshop or training on AI in medicine, leaving more than three-quarters reliant on self-directed or peer-mediated learning.

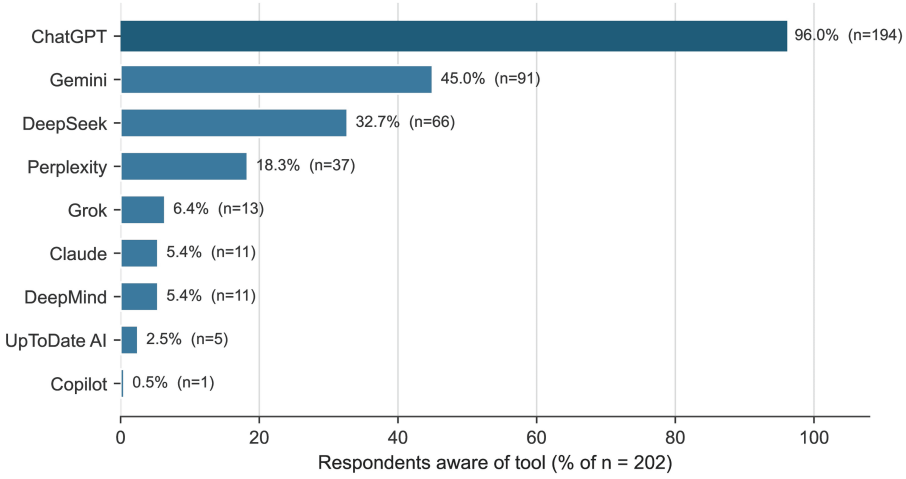


FIGURE 3: Awareness of specific generative AI tools

Proportion of respondents naming each tool when asked which generative AI tools they had heard of (n = 202); bars are labeled with the percentage and the frequency. The item allowed multiple responses, so proportions do not sum to 100%.

AI, artificial intelligence

Attitudes

Attitudes were broadly favorable but not uncritical (Figure 4). A large majority agreed that AI is useful for acquiring medical knowledge (165/202, 81.7%, 95% CI 75.8-86.4), and 129/202 (63.9%) supported integrating AI education into the curriculum, while opinion was divided on whether AI could replace traditional textbooks (yes, 78/202, 38.6%; no, 85/202, 42.1%; not sure, 39/202, 19.3%). Importantly, enthusiasm was tempered by caution: 138/201 (68.7%, 95% CI 61.9-74.7) expressed concern about ethical issues such as plagiarism and misinformation, and 148/202 (73.3%) would nonetheless recommend AI tools to peers.

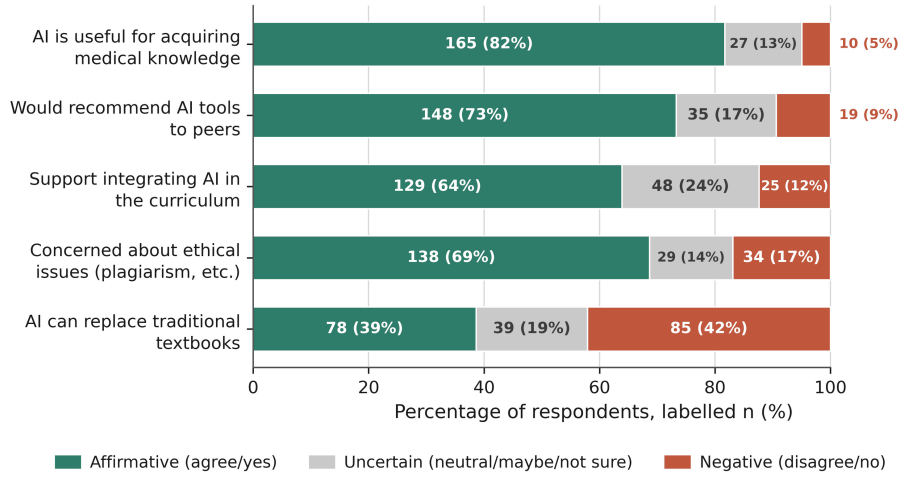


FIGURE 4: Attitudes toward AI in medical learning

For each item, bars show the proportion giving an affirmative, uncertain, or negative response (n = 201-202) and are labeled with the frequency and percentage, n (%). Where a segment is too narrow to hold its label, the label sits to the right of the bar.

AI, artificial intelligence

Practice

Use of AI for study was widespread: 173/202 (85.6%, 95% CI 80.1-89.8) reported having used AI tools to study or solve medical questions. Among those 173 users, 81 (46.8%) used AI daily or frequently. Twenty-

two respondents who reported never having used AI nonetheless answered the frequency item, which was intended for users only; all frequency analyses are therefore restricted to self-reported users, in both the numerator and denominator. The dominant purpose was summarizing topics (161/202, 79.7%), followed by writing assignments (110/202, 54.5%), answering clinical questions (92/202, 45.5%), and solving multiple-choice questions (87/202, 43.1%) (Figure 5). Use for answering clinical questions was almost identical in the two groups (UG 53/116, 45.7% vs PG 39/86, 45.3%).

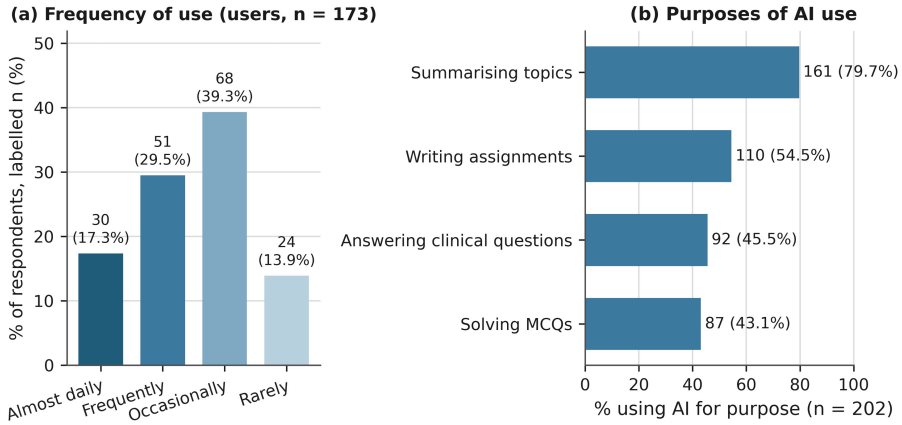


FIGURE 5: Practice patterns of AI use

(a) Frequency of use among the 173 respondents who reported ever using AI; respondents who reported never using AI are excluded from both numerator and denominator. (b) Purposes for which AI is used (n = 202); the item allowed multiple responses. Bars in both panels are labeled with the frequency and percentage, n (%).

MCQs, multiple-choice questions; AI, artificial intelligence

Undergraduate versus postgraduate comparison

The planned comparison between UG and PG trainees revealed broadly similar AI engagement, and no difference survived correction for multiple comparisons (Table 2; Figure 6). Differences in percentage points with 95% CIs are given for every indicator in Table 2. Awareness, ever-use, regular use, agreement that AI is useful, formal training, ethical concern, and curricular support did not differ significantly, with small effect sizes (all phi ⩽ 0.16), but the CIs are wide and do not exclude differences of 15-25 percentage points in either direction: formal training +4.9 points (-6.9 to +16.7) and ethical concern +11.5 points (-1.2 to +24.2), both higher among PGs. The single categorical item reaching nominal significance was willingness to recommend AI to peers, which was higher among PGs (UG 78/116, 67.2% vs PG 70/86, 81.4%; difference +14.2 points, 95% CI +2.3 to +26.0; P = 0.037); frequency of use among users did not differ on an ordinal Mann-Whitney test (P = 0.42). Under Benjamini-Hochberg correction across the family of 12 UG-PG comparisons, no result remained significant (smallest adjusted q = 0.44), so these signals are hypothesis-generating.

Indicator	UG % (n/N)	PG % (n/N)	Difference, PG-UG, percentage points (95% CI)	Test	P-value	q-value	φ
Aware of AI in medicine	93.1 (108/116)	97.6 (83/85)	+4.5 (-1.1 to +10.2)	Fisher	0.196	0.543	0.103
Ever used AI tools	87.9 (102/116)	82.6 (71/86)	-5.4 (-15.3 to +4.6)	χ^2	0.382	0.543	0.076
Regular use, daily or frequent (users only)	43.1 (44/102)	52.1 (37/71)	+9.0 (-6.1 to +24.1)	χ^2	0.313	0.543	0.088
Agree AI is useful	79.3 (92/116)	84.9 (73/86)	+5.6 (-5.0 to +16.1)	χ^2	0.407	0.543	0.071
Had formal AI training	20.7 (24/116)	25.6 (22/86)	+4.9 (-6.9 to +16.7)	χ^2	0.516	0.619	0.058
Concerned about ethics	63.8 (74/116)	75.3 (64/85)	+11.5 (-1.2 to +24.2)	χ^2	0.114	0.456	0.122
Support AI in curriculum	62.1 (72/116)	66.3 (57/86)	+4.2 (-9.1 to +17.5)	χ^2	0.640	0.698	0.043
Would recommend to peers	67.2 (78/116)	81.4 (70/86)	+14.2 (+2.3 to +26.0)	χ^2	0.037*	0.444	0.158
AI can replace textbooks	35.3 (41/116)	43.0 (37/86)	+7.7 (-5.9 to +21.3)	χ^2	0.336	0.543	0.078

TABLE 2: UG versus PG comparison of key indicators

Differences are PG minus UG in percentage points with 95% CIs. Regular use is restricted to the 173 respondents who reported ever using AI, in both numerator and denominator.

*Nominal P < 0.05; no comparison reached q < 0.05, so all are hypothesis-generating.

AI, artificial intelligence; UG, undergraduate; PG, postgraduate; χ^2, Pearson chi-square test with the Yates continuity correction; φ, phi/Cramér V effect size, computed from the uncorrected statistic; CI, confidence intervals; q, Benjamini-Hochberg-adjusted value across the twelve pre-planned comparisons

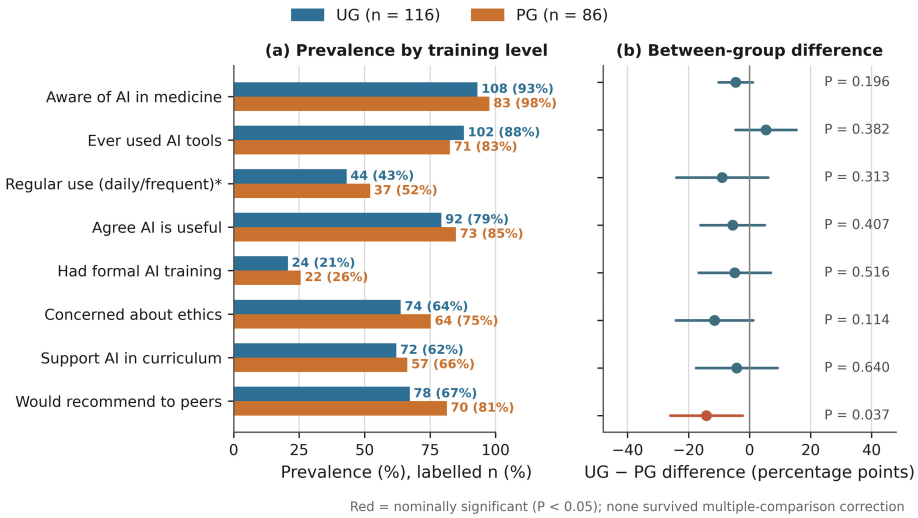


FIGURE 6: UG versus PG comparison

(a) Prevalence of each indicator by training level, labeled with the frequency and percentage, n (%). Denominators vary: an item left blank reduces it, and regular use is restricted to the respondents who reported ever using AI (UG 102, PG 71). (b) Between-group differences in percentage points with 95% CIs; only "recommend to peers" excludes zero, shown in red. No difference survived correction for multiple comparisons.

*Regular use is restricted to the 173 respondents who reported ever using AI (UG 44/102, PG 37/71); all other indicators use the full group denominators.

UG, undergraduate; PG, postgraduate; AI, artificial intelligence; CI, confidence intervals

To separate training level from the confounded age difference, the two nominal signals were reexamined with adjustment for age and gender (n = 201 for each model, representing respondents with complete age data). In these models, the PG advantages attenuated to nonsignificance: the higher likelihood of recommending AI was reduced after adjustment (crude odds ratio 2.13, 95% CI 1.09-4.15, P = 0.026; age- and gender-adjusted odds ratio 1.71, 95% CI 0.75-3.90, P = 0.20), and the PG attitude composite was similarly attenuated (adjusted β = +0.17, 95% CI -0.19 to +0.53, P = 0.35). Because age and training level are strongly correlated, these adjusted estimates should be read as a check on confounding rather than as evidence that training level is unimportant. In a post hoc model, ever-use was associated with age rather than level: each one-SD (approximately 4.5-year) increase in age roughly halved the odds of having used AI (adjusted odds ratio 0.45, 95% CI 0.28-0.72, P = 0.001), independent of training stage.

Composite KAP scores

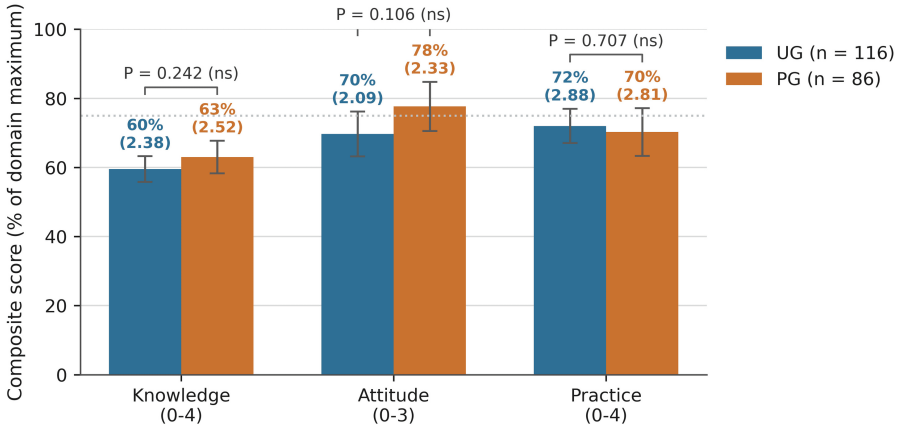
Composite domain scores were consistent with this picture (Table *3*; Figure *7*). Overall, respondents scored moderately on knowledge (2.44 ± 0.85; 61.0% of maximum) and more highly on the three-item attitude score (2.19 ± 1.05; 72.9%) and practice (2.85 ± 1.18; 71.3%). None of the three composites differed significantly between UG and PG groups; the attitude composite was higher among PGs but not significantly so (2.09 vs 2.33, t = -1.62, P = 0.106, d = -0.23). Awareness of ethical issues, which is reported as a separate item rather than included in the attitude composite, was also higher among PGs without reaching significance (63.8% vs 75.3%, P = 0.114).

Domain	Overall, mean ± SD (% of maximum)	UG	PG	P-value	q-value	d	α
Knowledge (0-4)	2.44 ± 0.85 (61.0%)	2.38 ± 0.82	2.52 ± 0.89	0.242	0.543	-0.17	0.43
Attitude (0-3)	2.19 ± 1.05 (72.9%)	2.09 ± 1.07	2.33 ± 1.01	0.106	0.456	-0.23	0.71
Practice (0-4)	2.85 ± 1.18 (71.3%)	2.88 ± 1.09	2.81 ± 1.31	0.707	0.707	0.06	0.70

TABLE 3: Composite knowledge, attitude, and practice scores by training level

Independent-samples t-test (Welch); The attitude composite contains the three enthusiasm items only; awareness of ethical issues is a caution rather than a positive attitude and is reported as a separate item in Table 2.

SD, standard deviation; UG, undergraduate; PG, postgraduate; d, Cohen d; α, Kuder-Richardson 20; q, Benjamini-Hochberg-adjusted value across the twelve pre-planned comparisons; no comparison reached q < 0.05.



Labels show the group mean as a percentage of the domain maximum, with the mean score itself in parentheses.

FIGURE 7: Composite knowledge, attitude, and practice scores by training level

Composite scores as a percentage of the domain maximum, with 95% CI. Knowledge and practice are scored 0-4 and attitude 0-3: awareness of ethical issues is a caution rather than an endorsement, so it is reported as a separate item in Table 2 rather than inside the attitude composite. Bars are labeled with the percentage of the domain maximum and, in parentheses, the group mean composite score. No composite differed significantly between the groups; the attitude difference (P = 0.106) did not reach nominal significance, and no comparison survived correction for multiple comparisons.

UG, undergraduate; PG, postgraduate; CI, confidence intervals

Sensitivity analyses

Because interns occupy an intermediate position, we repeated the principal comparisons with the 14 interns removed from the UG group (UG n = 102 vs PG n = 86). The overall conclusion of no robust between-group difference was unchanged. Two further sensitivity analyses were conducted in response to peer review. Excluding the three respondents who named another institution (n = 199) left the headline estimates essentially unchanged (ever-use 85.4%, formal training 23.1%, awareness 94.9%). Excluding, in addition, the single respondent who reported an age of 16 (n = 198) also changed little (ever-use 85.4%, formal training 23.2%, regular use among users 47.3%), and the attitude composite remained nonsignificant (2.08 vs 2.33, P = 0.099), while willingness to recommend remained nominally significant (P = 0.035). None survived multiplicity correction.

Discussion

The dominant finding is a gap. Generative AI use at this institution is widespread: 191/201 students (95.0%) were aware of AI in medicine, and 173/202 (85.6%) had used it for study, overwhelmingly through ChatGPT (194/202, 96.0%) and predominantly for summarizing material (161/202, 79.7%). Yet fewer than one in four (46/202, 22.8%) had received any structured instruction in these tools, and only 81 of the 173 users (46.8%) used them daily or frequently. Adoption is therefore largely informal, self-directed, and unsupervised. The

gap is not confined to low-stakes revision: 92/202 respondents (45.5%) already used AI to answer clinical questions, a proportion almost identical in the two groups (UG 45.7% vs PG 45.3%). The item did not define what a clinical question is, and because 84 of 116 UGs (72.4%) were in the first two preclinical years, for many respondents this likely means examination-style vignettes rather than decisions about a patient in front of them; the finding should be read accordingly. Two further findings frame the gap. Attitudes were positive but ethically tempered, with 138/201 (68.7%) expressing concern about plagiarism and misinformation and 129/202 (63.9%) supporting the integration of AI into the curriculum. No statistically robust difference was detected between UG and PG trainees on any indicator after correction for multiple comparisons.

That no robust stage difference emerged is what makes the gap tractable. The expectation that residents, exposed to the consequences of clinical error, would be markedly more cautious than junior students was not supported, although two of the three directional predictions pointed in the expected direction without reaching significance: PGs expressed more ethical concern (+11.5 points) rather than less. What ran against the expectation was willingness to recommend AI to peers, which was higher among PGs (+14.2 points), while measures of personal use did not differ. Neither signal was robust to correction for multiple testing, and both attenuated once age and sex were taken into account. A nonsignificant result must be interpreted with care, since absence of evidence is not evidence of absence [25]. Our study was powered only for differences of roughly 18-20 percentage points; no equivalence margin was specified in advance, and the CIs for several indicators remain compatible with differences of 15-25 points, which would matter for curriculum design. We therefore claim only that no statistically robust difference was detected, not that the two groups are equivalent. Because training level and age are strongly collinear (point-biserial $r = 0.58$), adjusting for age divides a largely shared effect between two correlated variables, so the attenuation seen after adjustment is expected and does not establish that training level is irrelevant; the adjusted models are best read as a check for confounding rather than as a decomposition. The post hoc association of lower ever-use with older age is consistent with a digital-native pattern rather than against it, and we note it only as a hypothesis for future work.

Our descriptive findings align closely with international KAP and perception studies: high awareness, broad use for academic tasks, and positive but ethically tempered attitudes are reported across diverse settings, from the United Kingdom and Europe to the Middle East and South Asia [12,21,26-29]. A 2026 global systematic review and meta-analysis using the KAP framework confirmed this pattern of high awareness, ethically cautious enthusiasm, and a persistent formal training gap across countries [30]. Studies focused on students' expectations of conversational LLMs report the same blend of enthusiasm and caution [31], and recent multi-item KAP surveys in Egypt, India, and the Gulf likewise found heavy adoption coexisting with limited formal training [18,20,29,32]. The two Indian studies that made the same within-institution UG-PG comparison disagree with each other, and our result sits between them. Gandhi et al., in a larger sample (404 UG and 201 PG), found PGs clearly ahead in AI readiness, whereas a tertiary-center survey from central India (188 UG and 129 PG) found KAP broadly similar across levels, as we did [14,15]. Three differences may explain why we detected no knowledge difference where Gandhi et al. did: their sample was three times larger and better powered for moderate differences; our UG group is weighted toward the first two preclinical years, which compresses the range of clinical exposure being compared; and our knowledge composite is short, with modest internal consistency, so it is insensitive to small gradations. Our specific contribution is to pair the size of the training gap with a direct within-institution comparison and to report the uncertainty around that comparison explicitly rather than treating a null result as equivalence.

Three implications follow, each anchored in our data. First, because exposure and the training gap are shared across levels, AI literacy should be embedded uniformly across the medical education continuum rather than targeted at a single cohort; a single institution-wide policy is both simpler and better matched to the data than stage-specific curricula and should make the distinction between consumer LLMs and curated clinical resources explicit. Second, this literacy must go beyond tool familiarity to the critical appraisal of AI output, including recognizing hallucinations and verifying information against trusted sources, because heavy, unsupervised use without these competencies risks overreliance and may blunt independent reasoning [5,33]; established AI literacy frameworks offer a ready scaffold [34]. This need is not hypothetical: 92/202 respondents (45.5%) already used AI for clinical questions while fewer than one in four had received any training, so verification skills should be taught before the clinical years. Third, the high level of ethical concern (138/201, 68.7%) is an opportunity rather than a barrier: clear institutional guidance on permissible use, disclosure, and academic integrity would meet a need that students themselves recognize [3].

Strengths of this study include a contemporary, post-generative-AI sample; a direct UG-PG comparison within a single institution using one instrument; and transparent reporting of denominators, CIs, effect sizes, and multiplicity. The UG group is not representative of UG training as a whole: 84 of 116 UGs (72.4%) were in the first or second year, with only nine third-year, nine final-year, and 14 intern respondents, so the comparison is largely between preclinical students and residents rather than clinical UGs and residents, and the stage contrast is correspondingly blunt. Against a sanctioned intake of 180 MBBS places per year and 259 PG places, the sample represents a small and self-selected fraction of those eligible. Several limitations temper interpretation. This is a single-center study using convenience sampling through digital channels, which oversamples the digitally engaged and probably inflates the absolute estimates of awareness and use; the training gap we report may therefore be a conservative estimate, since less digitally engaged students are

unlikely to have received more instruction. Three respondents named another institution and were retained in the primary analysis; excluding them changed nothing material. The between-group comparisons, being within-sample, are less affected by selection bias than the prevalence estimates. All data are self-reported, with attendant social desirability and recall bias, and one age entry is implausible and could not be verified. The instrument was investigator-developed and not externally validated; the knowledge composite in particular has low internal consistency (KR-20 = 0.43), so the composites are exploratory and the individual items in Table *1* and Table *2* carry more weight than the composite comparisons. Twenty-two nonusers answered an item intended for users, which suggests that some respondents read items inattentively. The cross-sectional design cannot establish causation or track rapidly evolving behavior, and the study was underpowered for small between-group differences. Faculty perspectives were not assessed. Multicenter and longitudinal studies with validated instruments are needed.

Conclusions

Generative AI has become a routine part of how medical students at this institution study, and a substantial minority already turn to it when answering clinical questions. Formal instruction has not kept pace. The distance between what students do with these tools and what they have been taught about them is the central problem identified by this study. We found no statistically robust difference between UG and PG trainees to suggest that this gap is concentrated in either group, although our study was not designed to demonstrate that the two groups are equivalent.

Because we found no evidence that the deficit is confined to one stage of training, there is no basis for targeting the remedy at a single cohort. AI literacy and ethics education should be embedded uniformly across the medical education continuum, supported by clear institutional guidance on permissible use, disclosure, and academic integrity. That teaching should extend beyond familiarity with the tools to the critical appraisal of their output, so that students learn to verify it against trusted sources before they reach the clinical years.

Appendices

Study questionnaire

Self-administered instrument delivered via Google Forms. The landing page presented the study aims and an informed-consent statement; only respondents who indicated agreement proceeded.

Section A - Socio-Demographic Information

1. Do you voluntarily agree to participate in this anonymous survey for academic purposes? (Yes / No)
2. Age: _____ years
3. Gender: Male / Female / Prefer not to say
4. Year of study: 1st year / 2nd year / 3rd year / Final year / Intern / Postgraduate
5. Institution name: _____

Section B - Knowledge

6. Have you heard of artificial intelligence (AI) in healthcare / medical education? (Yes / No)
7. Which AI tools have you heard of? (tick all that apply) ChatGPT / Google Gemini / DeepSeek / Perplexity / Google DeepMind / UpToDate AI / Claude / Grok / Copilot / Others (specify)
8. AI can help with: (tick all that apply) Learning and summarising topics / Writing assignments / Diagnostic assistance / Clinical decision-making / I don't know
9. Have you attended any workshop or training on AI in medicine? (Yes / No)

Section C - Attitude

10. Do you think AI is useful in acquiring medical knowledge? (Strongly agree / Agree / Neutral / Disagree / Strongly disagree)
11. Do you feel AI can replace traditional learning methods (like textbooks)? (Yes / No / Not sure)
12. Are you concerned about ethical issues in using AI tools for learning (e.g., plagiarism, misinformation)? (Yes / No / Not sure)
13. Do you support integrating AI education in the medical curriculum? (Yes / Maybe / No)

Section D - Practice

14. Have you ever used AI tools (like ChatGPT, UpToDate AI) to study or solve medical questions? (Yes / No)
15. If yes, how often do you use them? (Almost daily / Frequently / Occasionally / Rarely)
16. What do you use AI for? (tick all that apply) Summarising topics / Writing assignments / Answering clinical questions / Solving MCQs / I don't use AI
17. Would you recommend AI tools to your peers for academic help? (Yes / Maybe / No)
18. What are your suggestions for improving AI use in medical education? (open-ended)

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Asra Malik, Muzaffer Pukhta

Acquisition, analysis, or interpretation of data: Asra Malik

Drafting of the manuscript: Asra Malik

Critical review of the manuscript for important intellectual content: Muzaffer Pukhta

Supervision: Muzaffer Pukhta

Disclosures

Human subjects: Informed consent for treatment and open access publication was obtained or waived by all participants in this study. Institutional Review Board, Government Medical College Srinagar issued approval IRBGMC-SGR/Pharma/1492. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Acknowledgements

The authors thank the medical students of Government Medical College Srinagar who took part in the survey. During the preparation of this manuscript, the authors used Claude (Anthropic PBC, San Francisco, CA, USA) to assist with language editing, formatting the manuscript and reference list to journal style, and writing the Python code used to analyze the de-identified dataset. The study was conceived and designed, the data were collected, and the results were interpreted by the authors. All figures are charts plotted directly from the study data; no image was created or altered by a generative artificial intelligence tool. The authors checked every reference against its primary source, reviewed all artificial intelligence-assisted output, and take full responsibility for the content and accuracy of this article. The de-identified dataset and the study questionnaire are available from the corresponding author on reasonable request; the questionnaire is also reproduced in the appendices.

References

1. Sallam M: ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare (Basel). 2023, 11:887. 10.3390/healthcare11060887
2. Kung TH, Cheatham M, Medenilla A, et al.: Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digit Health. 2023, 2:e0000198. 10.1371/journal.pdig.0000198
3. Masters K: Ethical use of artificial intelligence in health professions education: AMEE Guide No. 158 . Med Teach. 2023, 45:574-84. 10.1080/0142159X.2023.2186203
4. Esteva A, Robicquet A, Ramsundar B, et al.: A guide to deep learning in healthcare . Nat Med. 2019, 25:24-9. 10.1038/s41591-018-0316-z
5. Gordon M, Daniel M, Ajiboye A, et al.: A scoping review of artificial intelligence in medical education: BEME Guide No. 84. Med Teach. 2024, 46:446-70. 10.1080/0142159X.2024.2314198
6. Bhattacharyya M, Miller VM, Bhattacharyya D, Miller LE: High rates of fabricated and inaccurate references in ChatGPT-generated medical content. Cureus. 2023, 15:e39238. 10.7759/cureus.39238
7. Grunhut J, Marques O, Wyatt AT: Needs, challenges, and applications of artificial intelligence in medical education curriculum. JMIR Med Educ. 2022, 8:e35587. 10.2196/35587
8. Charow R, Jeyakumar T, Younus S, et al.: Artificial intelligence education programs for health care professionals: scoping review. JMIR Med Educ. 2021, 7:e31043. 10.2196/31043
9. Ahmed Z, Bhinder KK, Tariq A, et al.: Knowledge, attitude, and practice of artificial intelligence among doctors and medical students in Pakistan: a cross-sectional online survey. Ann Med Surg (Lond). 2022, 76:103493. 10.1016/j.amsu.2022.103493
10. Swed S, Alibrahim H, Elkalagi NK, et al.: Knowledge, attitude, and practice of artificial intelligence among doctors and medical students in Syria: a cross-sectional online survey. Front Artif Intell. 2022, 5:1011524. 10.3389/frai.2022.1011524
11. Wartman SA, Combs CD: Medical education must move from the information age to the age of artificial intelligence. Acad Med. 2018, 93:1107-9. 10.1097/ACM.0000000000002044
12. Sit C, Srinivasan R, Amlani A, Muthuswamy K, Azam A, Monzon L, Poon DS: Attitudes and perceptions of

UK medical students towards artificial intelligence and radiology: a multicentre survey. Insights Imaging. 2020, 11:14. 10.1186/s13244-019-0830-7
13. Park GL, Beck Dallaghan GL, Bradley J, et al.: Generative artificial intelligence in medical training: utilization patterns across knowledge, patient care, systems reasoning, and innovation. Med Sci Educ. 2026, 36:5-10. 10.1007/s40670-025-02623-1
14. Gandhi R, Parmar A, Kagathara J, Lakkad D, Kakadiya J, Murugan Y: Bridging the artificial intelligence (AI) divide: do postgraduate medical students outshine undergraduate medical students in AI readiness?. Cureus. 2024, 16:e67288. 10.7759/cureus.67288
15. Prasad S, Khot A, Dakhale G: Knowledge, attitude and practice about artificial intelligence among medical undergraduate and postgraduate students in a tertiary care centre: a cross-sectional survey. Cureus. 2026, 18:e100576. 10.7759/cureus.100576
16. Civaner MM, Uncu Y, Bulut F, Chalil EG, Tatli A: Artificial intelligence in medical education: a cross-sectional needs assessment. BMC Med Educ. 2022, 22:772. 10.1186/s12909-022-03852-3
17. Mateen A, Kumar V, Singh AK, Yadav B, Mahto M, Hassan A, Nasir N: Impact of generative AI in medical education in India: a systematic review. Front Artif Intell. 2025, 8:1704785. 10.3389/frai.2025.1704785
18. Gopidesi M, Anil R, Mahesh S, Nagaraj C, Pichandi J: Bridging the artificial intelligence (AI) gap: a knowledge, attitude, and practice (KAP) study to advance medical education in rural Andhra Pradesh. Cureus. 2025, 17:e91822. 10.7759/cureus.91822
19. Sharma S, Baig VN, Saboo N, Kayat S: Artificial intelligence in healthcare: a cross-sectional study among students of a medical college in Jaipur, India. Perspect Clin Res. 2025, 16:227-8. 10.4103/picr.picr_145_24
20. Kumar A, Maurya A, Verma A, Binwal S, Kumar S, Saini A: Knowledge, attitude, and practice of artificial intelligence among doctors and medical students in medical institute of Northern India - a cross sectional study. Indian J Community Health. 2025, 37:1019-24. 10.47203/ijch.2025.v37i06.021
21. Jha N, Shankar PR, Al-Betar MA, Mukhia R, Hada K, Palaian S: Undergraduate medical students' and interns' knowledge and perception of artificial intelligence in medicine. Adv Med Educ Pract. 2022, 13:927-37. 10.2147/AMEP.S368519
22. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP: The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet. 2007, 370:1453-7. 10.1016/s0140-6736(07)61602-x
23. Eysenbach G: Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res. 2004, 6:e34. 10.2196/jmir.6.3.e34
24. World Medical Association: World Medical Association Declaration of Helsinki: ethical principles for Medical Research involving human participants. JAMA. 2025, 333:71-4. 10.1001/jama.2024.21972
25. Altman DG, Bland JM: Absence of evidence is not evidence of absence . BMJ. 1995, 311:485. 10.1136/bmj.311.7003.485
26. Pinto Dos Santos D, Giese D, Brodehl S, et al.: Medical students' attitude towards artificial intelligence: a multicentre survey. Eur Radiol. 2019, 29:1640-6. 10.1007/s00330-018-5601-1
27. Doumat G, Daher D, Ghanem NN, Khater B: Knowledge and attitudes of medical students in Lebanon toward artificial intelligence: a national survey study. Front Artif Intell. 2022, 5:1015418. 10.3389/frai.2022.1015418
28. Boillat T, Nawaz FA, Rivas H: Readiness to embrace artificial intelligence among medical doctors and students: questionnaire-based study. JMIR Med Educ. 2022, 8:e34973. 10.2196/34973
29. Almehmadi B, Bakhsh TM, Noor Elahi I, Hariri N: Awareness, perceptions, and opinions of artificial intelligence among undergraduate medical students at Umm al-Qura University, Saudi Arabia, in 2025: a cross-sectional study. Cureus. 2025, 17:e89741. 10.7759/cureus.89741
30. Visci P, Lupelli F, Calò F, Sirago G, Solarino B, Dell'Erba A, Ferorelli D: Undergraduate medical students and artificial intelligence: A global systematic review and meta-analysis using a knowledge-attitude-practice framework. Med Teach. 2026, 48:1791-803. 10.1080/0142159X.2026.2663855
31. Moldt JA, Festl-Wietek T, Madany Mamlouk A, Nieselt K, Fuhl W, Herrmann-Werner A: Chatbots for future docs: exploring medical students' attitudes and knowledge towards artificial intelligence and medical chatbots. Med Educ Online. 2023, 28:2182659. 10.1080/10872981.2023.2182659
32. Ghanem OA, Hagag AM, Kormod ME, et al.: Medical students' knowledge, attitudes, and practices toward generative artificial intelligence in Egypt 2024: a Cross-Sectional study. BMC Med Educ. 2025, 25:790. 10.1186/s12909-025-07329-x
33. Zhai C, Wibowo S, Li LD: The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review. Smart Learn Environ. 2024, 11:28. 10.1186/s40561-024-00316-7
34. Ng DTK, Leung JKL, Chu SKW, Qiao MS: Conceptualizing AI literacy: an exploratory review. Comput Educ Artif Intell. 2021, 2:100041. 10.1016/j.caeai.2021.100041