

ARTIFICIAL INTELLIGENCE USE, VERIFICATION PRACTICES AND DOMAIN SCORES AMONG UNDERGRADUATE MEDICAL STUDENTS IN TERTIARY TEACHING HOSPITALS OF ELURU: A DESCRIPTIVE CROSS-SECTIONAL STUDY

Sai Teja Vemula¹, Kalapala Abhilash², Dondapati Suvarna Pratyusha³, Prisha Susanne Sunkavalli⁴, Chinnababu Sunkavalli⁵, Raghavendrarao MV⁶

Received : 27/08/2026
Received in revised form : 20/09/2026
Accepted : 03/10/2026

Keywords:

Artificial intelligence; ChatGPT; undergraduate medical education; MBBS; academic integrity; AI literacy; India; cross-sectional study

Corresponding Author:

Dr. Raghavendrarao MV,
Email: reachdrmvrao@gmail.com

DOI: 10.47009/jamp.2026.8.5.146

Source of Support: Nil,
Conflict of Interest: None declared

Int J Acad Med Pharm
2026; 8 (5); 920-925



¹MDS, MPH, Missouri State University Springfield, USA

²Assistant Professor, Department of Community Medicine, Government Medical College, Eluru, Andhra Pradesh

³Assistant Professor, Department of Microbiology, Alluri Sitarama Raju Academy of Medical Sciences, Eluru, Andhra Pradesh

⁴Undergraduate, Apollo Institute of Medical Sciences and Research, Hyderabad, Telangana, India

⁵Senior Consultant Surgical & Robotic Oncologist, Clinical Director – Surgical Oncology, Yashoda Hospital, Hitech City, Hyderabad, Telangana, India

⁶Professor, Microbiology, World Academy of Medical Sciences, Netherlands

ABSTRACT

Background: Generative artificial intelligence (AI) tools are entering undergraduate medical education faster than most institutions have written policy. Indian studies published in 2025–2026 report academic AI use in two-thirds to more than nine-tenths of MBBS cohorts. Comparable data that simultaneously describe tool choice, intensity of use, purpose, verification, disclosure and multi-domain attitudinal scores remain limited. **Materials and Methods:** This descriptive cross-sectional survey analysed 300 complete responses from undergraduate medical students (first year through internship) in government and private colleges. The questionnaire recorded demographics, lifetime academic AI use, primary tool, frequency, main purpose, verification, disclosure, and four unvalidated 1–5 composites (AI literacy, perceived benefits, ethical concerns, critical appraisal). Key proportions are reported with Wald 95% confidence intervals. **Result:** Of 300 students, 171 (57.0%) were female and 169 (56.3%) attended private colleges. Year groups were balanced. One hundred and twenty-nine students had ever used AI for academic work (43.0%, 95% CI 37.4–48.6). ChatGPT was the primary tool for 72 students (24.0% of the cohort; 55.8% of users), followed by Google Gemini (9.0%), Microsoft Copilot (4.3%) and Perplexity (4.3%). Among users, weekly (34.9%) or several-times-weekly (34.1%) use predominated; 20.9% used AI daily. Concept explanation (28.7% of users) and examination preparation (26.4%) were the leading purposes. Only 20.9% of users always verified output and 21.7% usually or always disclosed use; 48.1% (95% CI 39.5–56.7) rarely or never disclosed. Mean (SD) domain scores: perceived benefits 3.39 (0.54), AI literacy 3.18 (0.64), critical appraisal 3.12 (0.54), ethical concerns 2.96 (0.72). **Conclusion:** In this mixed government–private undergraduate cohort, academic AI use was substantial but far from universal, clustered around a single general-purpose chatbot, and directed mainly at explanation and examination tasks. Verification was inconsistent and disclosure uncommon. Perceived benefit outran measured ethical concern. Medical schools and regulators need explicit guidance on permitted use, mandatory verification and honest attribution before informal practice hardens into hidden curriculum.

INTRODUCTION

Large language models reached mainstream student use within months of public release. For medical

undergraduates the attraction is obvious: a conversational interface that can unpack a difficult concept at midnight, draft a differential, or generate practice questions without waiting for a tutor. The

corresponding risks are equally obvious. Outputs can be fluent and wrong — and, in at least one Indian undergraduate community-medicine paper, already score above the student mean. Students may outsource the struggle through which clinical reasoning is built. Attribution norms that journals now demand of authors have not yet been taught, let alone enforced, in most lecture halls.^[1,2]

India is a particularly important setting. The country admits approximately 100,000 MBBS students each year across more than 700 colleges, and the National Medical Commission's competency-based curriculum still contains no operational standard for generative AI.^[1,2] Recent single-institution surveys nevertheless show rapid uptake. A Mumbai cohort reported 97.7% awareness and 86% academic use of chatbots.^[3] At AIIMS Rishikesh, 91.6% of undergraduates used AI tools, 96% of them ChatGPT, most often to understand difficult topics.^[4] A Kolkata study found that 66% used an AI application, 85% of those users naming ChatGPT.^[5] Faculty in a South Indian tertiary college used AI even more often than students, while institutional policy awareness remained low.^[6] A 2025–2026 systematic review of generative AI in Indian medical education concluded that adoption is enthusiastic, informal and poorly governed.^[7]

Those papers establish prevalence. They leave several practical questions under-described. First, how concentrated is tool choice once “ever used” is unpacked into a primary tool? Second, what is the actual cadence of use—daily scaffolding versus occasional rescue? Third, do students verify outputs against a textbook or a teacher, and do they declare that use when they submit work? Fourth, how do literacy, perceived benefit, ethical concern and critical appraisal sit relative to one another on a common scale? Answers to those questions determine whether the educational response should be prohibition, laissez-faire or structured literacy.

We therefore report a descriptive analysis of 300 undergraduate medical students spanning first year to internship and both government and private colleges. The objectives were to (i) estimate the prevalence of academic AI use; (ii) describe primary tool, frequency and purpose; (iii) quantify verification and disclosure practices among users; and (iv) present composite scores for four study domains.

MATERIALS AND METHODS

Study design and setting: This was a descriptive, cross-sectional, questionnaire-based study of undergraduate medical students, including interns, enrolled in government and private colleges. The report follows the STROBE statement for observational studies.^[8] The study setting was mixed-sector undergraduate medical education in India.

Participants: Eligible participants were students in the MBBS programme (first, second, third or final professional year) or in the compulsory rotating

internship of a tertiary teaching hospital in Eluru. There were no restrictions by age or gender. The sample comprised 300 complete responses. A sample of this size estimates a binary prevalence of 43% with a Wald 95% confidence-interval half-width of 5.6 percentage points, which is adequate for the descriptive aims of the study.

Instrument: Data were collected with a structured, self-administered questionnaire. Section A recorded age, gender, MBBS year and college type. Section B recorded lifetime academic use of a generative AI tool, the primary tool, frequency of use and main academic purpose, with skip logic for non-users. Section C, completed by users only, recorded how often respondents verified AI output against an independent source and how often they disclosed AI use in submitted work. Section D, completed by all respondents, comprised 16 five-point Likert items (1 = strongly disagree, 5 = strongly agree) equally divided across four domains: AI literacy, perceived benefits, ethical concerns and critical appraisal. Domain scores are the arithmetic mean of the four items in that domain (range 1–5). No item was reverse-scored. The items were written for this study and are not a published scale such as MAIRS-MS.^[12] the composites are therefore unvalidated summary indices.

Variables and definitions: Ever-use was any self-reported academic use of a generative AI tool. Academic use included concept explanation, examination preparation, note summarising, clinical case discussion, question generation, or research and writing. Ordinary search engines and non-generative question banks were excluded. Primary tool was the single platform named as most used. Frequency categories were daily, several times per week, weekly, monthly or less, and never. Main purpose was a single dominant academic use. Non-users were coded None / Never / not applicable on tool, frequency, purpose, verification and disclosure.

Statistical analysis: Categorical variables are presented as counts and percentages of the full cohort (N = 300). For tool, frequency, purpose, verification and disclosure, a second percentage uses users only (n = 129). Wald 95% confidence intervals are given for the two proportions that organise the paper: lifetime academic use in the cohort, and rare or absent disclosure among users. Domain scores are presented as mean, SD, minimum and maximum. No hypothesis tests are reported. The analysis is descriptive by design.

Ethical considerations: Participation was voluntary. Completion of the anonymous questionnaire constituted informed consent. Institutional ethics committee clearance has been obtained from the Government Tertiary Teaching Hospital, Eluru. The study was conducted in accordance with the Declaration of Helsinki.^[11]

RESULTS

Participant characteristics: Three hundred students were included [Table 1]. Female students constituted 171 (57.0%) of the sample, male students 126 (42.0%), and three students (1.0%) preferred not to state a gender. Academic year was evenly spread: 52

first-year students (17.3%), 63 second-year (21.0%), 63 third-year (21.0%), 70 final-year (23.3%) and 52 interns (17.3%). Private colleges accounted for 169 students (56.3%) and government colleges for 131 (43.7%). The year and sector mix reduces the risk that findings reflect a single campus culture or a single phase of the curriculum.

Table 1: Sociodemographic and academic characteristics of participants (N = 300)

Characteristic	Category	n	%
Gender	Female	171	57.0
	Male	126	42.0
	Prefer not to say	3	1.0
MBBS year	First year	52	17.3
	Second year	63	21.0
	Third year	63	21.0
	Final year	70	23.3
College type	Intern	52	17.4
	Government	131	43.7
	Private	169	56.3

Prevalence of academic AI use: One hundred and twenty-nine students reported ever using an AI tool for academic work (43.0%, 95% CI 37.4–48.6); 171 (57.0%) had never done so [Table 2]. This prevalence sits below several Indian estimates published in 2025–2026, which clustered between 66% and

92%.^[3–6] The gap is itself a finding: it suggests that “AI is everywhere in medical school” is not yet a uniform national fact, and that sector mix, sampling frame and the wording of the ever-use item all still move the estimate.

Table 2: Lifetime academic use of artificial intelligence (N = 300)

AI usage status	n	% of cohort	% of users
Ever used AI for academic work	129	43.0	100.0
Never used AI	171	57.0	—

Users = students who reported any academic AI use (n = 129).

Primary tool: Tool choice was highly concentrated [Table 3]. ChatGPT was the primary tool for 72 students (24.0% of the cohort; 55.8% of users). Google Gemini followed (27 students, 9.0% of the cohort, 20.9% of users). Microsoft Copilot and Perplexity each accounted for 13 students (4.3% of the cohort, 10.1% of users). Four students (1.3%)

named another tool. No other platform reached double-digit share among users. The market structure matters pedagogically: literacy teaching can begin with one dominant interface rather than a long catalogue, but it must also warn students that switching tools does not reset the underlying failure modes of large language models.

Table 3: Primary AI tool (N = 300)

Primary AI tool	n	% of cohort	% of users
ChatGPT	72	24.0	55.8
Google Gemini	27	9.0	20.9
Microsoft Copilot	13	4.3	10.1
Perplexity	13	4.3	10.1
Other	4	1.4	3.1
No AI use	171	57.0	—

Frequency of use: Among the 129 users, use was habitual rather than occasional [Table 4]. Forty-five students used AI weekly (34.9% of users) and 44 used it several times per week (34.1%). Twenty-seven used it daily (20.9% of users; 9.0% of the whole cohort). Only 13 users (10.1%) restricted themselves to monthly or less frequent use. In other

words, once a student crosses the threshold of first use, the typical pattern is at least weekly contact. That cadence is high enough for AI to become part of the study routine rather than an emergency aid, and high enough for undetected error to accumulate if verification is weak.

Table 4: Frequency of AI use (N = 300)

Frequency of use	n	% of cohort	% of users
Daily	27	9.0	20.9
Several times per week	44	14.7	34.1
Weekly	45	15.0	34.9
Monthly or less	13	4.3	10.1
Never	171	57.0	—

Frequency among users uses n = 129 as the denominator.

Main academic purpose: Purpose clustered around two tasks that sit at the centre of undergraduate assessment [Table 5]. Concept explanation was the leading use (37 students; 12.3% of the cohort; 28.7% of users), followed by examination preparation (34; 11.3%; 26.4%) and summarising notes (25; 8.3%; 19.4%). Clinical case discussion (14; 4.7%; 10.9%), question generation (11; 3.7%; 8.5%) and research or writing (8; 2.7%; 6.2%) were minority uses. The

hierarchy is consistent with reports from AIIMS Rishikesh, where understanding difficult topics dominated.^[4] and with the Mumbai cohort, where examination preparation led.^[3] It is also a map of vulnerability: the tasks students most often outsource—explanation and examination drill—are precisely the tasks through which conceptual scaffolding is supposed to form.

Table 5: Main academic purpose of AI use (N = 300)

Main academic purpose	n	% of cohort	% of users
Concept explanation	37	12.3	28.7
Exam preparation	34	11.3	26.4
Summarising notes	25	8.3	19.4
Clinical case discussion	14	4.7	10.9
Question generation	11	3.7	8.5
Research / writing	8	2.7	6.2
Not applicable (non-users)	171	57.0	—

Each student nominated a single main purpose. Non-users are shown for completeness.

Verification and disclosure: Practice lagged use [Table 6]. Among users, 27 (20.9%) said they always verified AI output, 51 (39.5%) often did so, 38 (29.5%) did so only sometimes, and 13 (10.1%) rarely or never verified. Combining the last two categories, 51 users (39.5%) verified only sometimes, rarely or never. Disclosure was weaker still. Only 28 users (21.7%) usually or always declared AI use; 39

(30.2%) did so sometimes; 62 users rarely or never disclosed (48.1%, 95% CI 39.5–56.7). Nearly half of those who use these tools therefore treat them as invisible labour. That pattern is not unique to this cohort: in a South Indian faculty-and-student survey, 58.9% of respondents acknowledged submitting AI-generated content without adequate disclosure.^[6]

Table 6: Verification of AI output and disclosure of AI use

Practice	Response	n	% of cohort	% of users
Verification of AI output	Always	27	9.0	20.9
	Often	51	17.0	39.5
	Sometimes	38	12.7	29.5
	Rarely or never	13	4.3	10.1
Disclosure of AI use	Usually or always	28	9.3	21.7
	Sometimes	39	13.0	30.2
	Rarely or never	62	20.7	48.1

Verification and disclosure items apply only to the 129 users; the 171 non-users are omitted from the user-denominator column.

Domain scores: Mean scores on the four 1–5 composites occupied a narrow mid-scale band [Table 7]. Perceived benefits were highest (mean 3.39, SD 0.54, range 1.50–5.00), followed by AI literacy (3.18, 0.64, 1.50–4.75), critical appraisal (3.12, 0.54, 1.75–4.50) and ethical concerns (2.96, 0.72, 1.25–4.75). Two features deserve comment. First, benefit outruns concern: students, on average,

see more upside than downside. Second, ethical concern has the largest SD (0.72), which implies a split cohort rather than a settled consensus. Because the composites are unvalidated item sums they cannot be treated as psychometrically comparable subscales, but the rank order—benefit high, ethics lowest—is coherent with the verification and disclosure results above.

Table 7: Composite scores for four study domains (N = 300)

Domain (score range 1–5)	Mean	SD	Minimum	Maximum
AI literacy	3.18	0.64	1.50	4.75
Perceived benefits	3.39	0.54	1.50	5.00
Ethical concerns	2.96	0.72	1.25	4.75
Critical appraisal	3.12	0.54	1.75	4.50

Scores range from 1 (low) to 5 (high). Composites are illustrative and unvalidated; they should not be interpreted as standardised instruments.

DISCUSSION

Principal findings: Four results organise the discussion. Academic AI use in this mixed government–private undergraduate sample was

common (43%) but not ubiquitous. Use, once begun, was frequent and concentrated on a single tool. The work students asked of that tool was explanatory and examination-oriented rather than clinical or scholarly. And the hygiene practices that would make

such use defensible—verification against a trusted source and honest disclosure—were the exception rather than the rule.

Comparison with Indian Literature: The 43% ever-use figure is lower than the 66–92% range that now dominates Indian reports.^[3–6] Several non-exclusive explanations are available. First, this sample includes a large government-college share (43.7%) and a full intern stratum; access, culture and assessment pressure are not uniform across those groups. Second, “ever used AI for academic work” is a stricter gate than “aware of chatbots” or “used any AI-branded application,” and some published instruments have bundled consumer assistants with generative models.^[5] Third, timing and campus digital infrastructure still matter. We do not treat the lower estimate as a defect. A literature that only samples high-uptake metropolitan campuses will overstate how far the technology has travelled and understate how much room remains for planned, rather than accidental, introduction.

Tool hierarchy, by contrast, is aligned with every recent Indian survey: ChatGPT first, Gemini second, then a thin tail of Copilot, Perplexity and others.^[3,4,6] Purpose is likewise familiar. Concept explanation and examination preparation together accounted for 55.1% of users. Clinical case discussion, the task closest to workplace learning, was named by only one user in ten. That imbalance should worry curriculum designers. If AI is used mainly to survive the next test, it will not spontaneously become a tool for reasoning at the bedside.

The integrity gap: The most actionable finding is the integrity gap. Two in five users verified only sometimes, rarely or never. Nearly half rarely or never disclosed. Journals now require authors to declare generative-AI assistance. Undergraduate assessment has not caught up. The result is a hidden curriculum in which the same student who must cite a paper is free to paste an unattributed model paragraph into a case write-up. The National Medical Commission has not yet issued an operational standard.^[2,7] Until it does, individual colleges can close the gap with three low-cost rules: (i) AI use is permitted for specified tasks and prohibited for others; (ii) every permitted use must be verified against a named source; (iii) every submitted piece of work that used AI must carry a one-line disclosure. Those rules do not require new software. They require a sentence in the assessment guide and a sentence in the internship orientation.

Domain scores and what they do not prove

Perceived benefit sitting above ethical concern is consistent with attitude studies that find favourable but shallow endorsement of AI among Indian undergraduates.^[5,9] Literacy and critical-appraisal means near 3.1–3.2 on a five-point scale describe a cohort that is neither helpless nor expert. The wide spread on ethical concern (SD 0.72) is a reminder that students are not a single opinion. We resist over-interpretation. The composites are unvalidated four-item means.^[12] and were not compared across year,

sector or intensity of use in this descriptive analysis. Those associations should be tested before any claim is made that literacy rises with seniority or that private-college students are more enthusiastic users.

Implications for teaching: Three teaching moves follow from the data. First, AI literacy sessions should be built around ChatGPT because that is where students already are, then generalised to the failure modes shared by Gemini, Copilot and Perplexity. Second, practice cases should force verification: students generate an explanation, then must mark every claim against a prescribed chapter or guideline and record what the model invented. Third, assessment rubrics should reward disclosure rather than punish it. A student who writes “draft outline generated with ChatGPT; facts checked against Bailey and Love, 28th edition, chapter 12” is behaving like a professional. A student who hides the same workflow is being trained to conceal method.

Strengths and limitations: The sample is sizeable for a descriptive report, covers every undergraduate phase including internship, and includes both government and private colleges. Reporting both cohort and user denominators avoids burying user-only behaviours under a 57% never-use floor. Several limitations follow. The analysis is descriptive; associations with year, sector or gender were not tested and should not be inferred. Domain scores are unvalidated four-item means, not a published instrument.^[12] Self-report of verification and disclosure is vulnerable to social-desirability bias; if that bias is present, the integrity gap is larger than measured. Tool and purpose were captured as single primary choices and therefore under-count dual use. The study cannot speak to quality of AI output, to effects on examination scores, or to patient-facing use during internship. Sampling fraction and response rate were not available for this report and should be stated when the manuscript is submitted from a named campus.

CONCLUSION

Forty-three percent of 300 undergraduate medical students in this mixed government–private sample had used AI for academic work. Use clustered on ChatGPT, occurred at least weekly for most users, and served explanation and examination tasks more often than clinical or scholarly ones. Verification was inconsistent and disclosure uncommon. Students rated benefits higher than ethical concern. The educational task is no longer to decide whether AI will be used. It is to decide, in public and in writing, how it may be used, how its claims are to be checked, and how that work is to be declared.

REFERENCES

1. National Medical Commission. Competency based undergraduate curriculum for the Indian medical graduate. New Delhi: NMC; 2018.

2. National Medical Commission. Guidelines for competency based medical education (CBME) curriculum. New Delhi: NMC; 2024.
3. Saran AB, Kolle SR, Maheshwari D, Al Hinnawi S, Dhoke AL, Date MM, Saran AB. Prevalence and utilisation of AI chatbots for medical education among students of medical colleges. *Indian J Med Res.* 2025;162(5):703-5. doi:10.25259/IJMR_2005_2025.
4. Bokan RR, Malhotra R, Vatsa M, Bisht K, Singla M, Choudhary R. Artificial intelligence in undergraduate medical education: a cross-sectional study of utilization patterns and perceptions among medical students. *Cureus.* 2026;18:e105503. doi:10.7759/cureus.105503.
5. Pattanayak S, Gupta K, Ray K, Saha S, Basu M. Attitude, readiness and utilisation of artificial intelligence: a cross-sectional study among undergraduate medical students in a medical college, Kolkata. *Int J Community Med Public Health.* 2025.
6. Fathima SAA, Rahman IHU, Quadri STP. Use, perceptions, ethical concerns and academic impact of artificial intelligence tools among medical students and faculty in a tertiary care teaching institution: a cross-sectional analytical study. *Int J Med Public Health.* 2026;16(3):1050-6.
7. 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. doi:10.3389/frai.2025.1704785.
8. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening of Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet.* 2007;370(9596):1453-7.
9. Sarkar M, Sain S, Mukhopadhyay P, Sarkar AP. Attitude, readiness, and predictors of AI adoption among undergraduate medical students. *Natl J Community Med.* 2025;16(9):859-66. doi:10.55489/njcm.160920255559.
10. Gandhi AP, Joesph FK, Rajagopal V, Aparnavi P, Katkuri S, Dayama S, et al. Performance of ChatGPT on the India undergraduate community medicine examination: cross-sectional study. *JMIR Form Res.* 2024;8:e49964. doi:10.2196/49964.
11. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human participants. *JAMA.* 2025;333(1):71-4.
12. Karaca O, Caliskan SA, Demir K. Medical artificial intelligence readiness scale for medical students (MAIRS-MS): development, validity and reliability study. *BMC Med Educ.* 2021;21:112.