JOURNAL OF CME
2025, VOL. 14, 2437293
https://doi.org/10.1080/28338073.2024.2437293

EUROPEAN CME Forum
Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

Artificial Intelligence and ChatGPT in Medical Education: A Cross-Sectional Questionnaire on students' Competence

L. Maaß[a], C. Grab-Kroll [b], J. Koerner [b], W. Öchsner [b,c], M. Schön [a], DAC Messerer[d], TM Böckers [a] and Anja Böckers [a]

[a]Institute for Anatomy and Cell Biology, Faculty of Medicine, Ulm University, Ulm, Germany; [b]Office of the Dean of Studies, Faculty of Medicine, Ulm University, Ulm, Germany; [c]Department of Anesthesiology and Intensive Care Medicine, University Hospital Ulm, Ulm, Germany; [d]Institute of Transfusion Medicine, University of Ulm, Ulm, Germany

ABSTRACT

Artificial intelligence is rapidly transforming the field of health science and medical education, but less is known about the students´ competencies related to knowledge, skills and attitudes towards the application of AI tools like ChatGPT. Therefore, a unicentric questionnaire-based cross-sectional study was applied to students in the medical field ($n = 207$). The data revealed that while most students were familiar with ChatGPT (66.7%), other AI tools were significantly less known or utilised for study purposes. Students approached AI tools rather informally, often preferring to use them as a simple search engine. More than half of the students admitted that they were not sufficiently informed about the underlying technology of AI. They applied ChatGPT in a self-directed manner but expressed considerable uncertainty regarding effective prompt engineering and ChatGPT's legal implications. Overall, the majority of respondents showed interest in and positivity towards the introduction of AI. However, they did not feel adequately prepared to handle AI confidently, leading many to express interest in further training. This training should be directly related to students' professional roles, e.g. as a physician. The three most favoured AI-topics for voluntary learning formats were AI in their studies (62.5%), AI in general (58.0%), and the use of AI in scientific writing (57.0%). Notable subgroup differences related to the students" gender or self-assessed study performance were observed and should be considered in future research.

ARTICLE HISTORY
Received 25 July 2024
Revised 16 November 2024
Accepted 27 November 2024

KEYWORDS
Artificial intelligence; ChatGPT; medical education; health education; attitudes; digital competence; AI literacy

Introduction

Artificial intelligence (AI) is rapidly transforming the field of health science and medical education. One of the most exciting developments in recent years has been the emergence of Large Language Models (LLM), such as ChatGPT which debuted in 2022. Since ChatGPT has demonstrated its capability to pass the United States Medical Licensing Examination (USMLE) [1,2], it has become increasingly clear that these AI tools significantly impact medical education, as evidenced by an increasing body of literature related to ChatGPT [3]. ChatGPT is poised to be a game changer in health care and medical education, aiding students and educators in approaching learning and teaching in innovative ways. However, it also raises ethical and legal concerns, such as issues related to inaccurate content, data privacy, and risk of plagiarism [4–6]. Several authors detail the benefits of AI, including its potential to facilitate adaptive learning, provide individual feedback and tutoring, reduce administrative workload, develop new assessment methods, brainstorm ideas, and support scientific research activities. However, they also highlight challenges, such as data security, threats to academic integrity due to potential biases or inaccuracies and the risks associated with excessive reliance on AI or unfair assessment practices [7–9].

A fundamental milestone in AI was the development of machine learning (ML) in the 1960s [10] which involved analysing training data and to create algorithms applicable to new data. Basic research in chatbots followed a few years later [11]. In 1964, ELIZA was introduced, marking the first use of a rudimentary form of natural language processing (NLP) to enable simple conversations between humans and computers in a medical context [12,13]. However,

CONTACT Anja Böckers anja.boeckers@uni-ulm.de Institute for Anatomy and Cell Biology Ulm University Faculty of Medicine Albert-Einstein-Allee 11 89081 Ulm, Germany

Supplemental data for this article can be accessed online at https://doi.org/10.1080/28338073.2024.2437293

it took several decades of AI development before a Chatbot like ChatGPT 3.5 (Chat Generative Pre-trained Transformer) was released by OpenAI Inc. San Fransisco, USA (https://chat.openai.com) in November 2022. ChatGPT is an LLM trained on billions of words taken from books, articles, and other internet content up until 2021. To process this data, neural networks (NN) were employed to generate new natural language content on demand [14]. As a result, NNs learned to combine words naturally and perform language tasks through complex algorithms that are ultimately so intricate that they are no longer fully reproducible, even by the developers themselves (AI blackbox). ChatGPT is an AI-based LLM that can process and generate human language. It is available free of charge for a variety of daily tasks, including text translation, writing, answering, and providing explanations. Therefore, just two months after its release, it became the fastest-growing consumer application, reaching about 100 million monthly users [15].

LLMs like ChatGPT have the potential to revolutionise health sciences and medical education in several ways. They could enhance the efficiency and precision of medical research by searching large datasets, generating literature reviews, identifying patterns and trends that would be difficult or impossible to find manually, and even writing academic abstracts or manuscripts [6,15–18].

In the area of patient care, AI tools are being used to support physicians in their day-to-day tasks and decision-making processes related to diagnosis [19] or treatment [20]. Several applications in various medical fields have already received official certification from the Food & Drug Administration (FDA) [21], demonstrating their relevance, particularly in image analysis technique for radiology, ophthalmology, or dermatology. New applications are being introduced to the market daily. The ability to analyse large amounts of data collected in intensive care settings or from wearable devices like smartwatches will aid in the early detection of conditions like renal failure or of cardiovascular diseases [22,23].

In particular, LLMs like ChatGPT can help to overcome language barriers between healthcare professionals and patients by translating medical terminology into plain English or answering routine patient queries [24]. They can also facilitate daily administrative tasks, such as writing discharge summaries, and provide patients with medical information and digital therapeutic support [18,25]. However, current AI-tools are not the best solution for every task. Recent studies summarised by Li et al. in 2023 showed inconsistent results regarding ChatGPT's efficiency. Consequently, the authors concluded that the then-current version of ChatGPT was somewhat unreliable for clinical use, recommending instead LLM applications specifically trained on biomedical data [26].

ChatGPT and AI are also poised to change the landscape of medical education, not only in clinical decision-making [26–30]. Researchers confirm the potential of ChatGPT for medical training as it allows students to access current information at any time, to support individual and interactive learning, to improve communication skills and therefore might help to educate more competent doctors. In detail, ChatGPT might be used like a dictionary [31], enhance clinical reasoning [32], or serve as an assessment tool to create multiple-choice questions (MCQ) [33]. LLMs may also generate feedback on student work, helping them improve their skills [28–31]. Surprinsingly, it could even be a tool to teach empathy, as its responses have been rated as more empathetic than those of the clinicians [34]. However, researchers in the medical educational field caution that ChatGPT is prone to providing incorrect information [35] and they highlight the need to address ethical issues such as data bias, user consent, data ownership, and data privacy [36]. Following the release of ChatGPT and its impact, there is a growing call for clearer guidelines that promote digital progress while maintaining existing standards [37].

Future physicians must learn to apply these new technologies in their work [38]. Therefore, teachers in higher education are tasked with teaching AI-related competencies. The European Commission has previously published a Framework for the Digital Competence of Educators [39] and has begun to include specifications to address the requirements of new digital technology like AI. Schönbächler identified "information literacy" as a core competence [40], while Laupichler et al., 2022, introduced the similar term "AI literacy" [41]. However, AI literacy is not well defined and further research is needed to clarify which knowledge, skills, and attitudes related to AI should be incorporated into higher education.

But how can AI literacy and its progress be measured among students or teachers? To address this, Karaca et al., 2021, developed a valid and reliable psychometric instrument called MAIRS-MS to assess readiness for medical AI among medical students [42]. This instrument was translated into German shortly afterwards by Laupichler et al. in 2022 [43]. The authors propose that "*medical artificial intelligence readiness is a healthcare provider's preparedness state in knowledge, skills, and attitude to utilize health-care AI application*". Another readiness questionnaire was developed and administered to multinational medical students and doctors [44]. The findings revealed a low level of familiarity with AI

in both groups, leading to the conclusion that physicians and medical students need to understand the key concepts behind AI, emphasising that AI must be integrated into medical training. When asked, 96.2% of medical students expressed a desire to be trained in knowledge and skills related to AI applications. However, only 6% felt capable to adequately informing a patient about risks and benefits of an AI-application [45].

In their review, Lee et al. provided several recommendations regarding curricular objectives and listed U.S. medical schools that have implemented pilot AI curricula [31]. They also identified challenges associated with introducing AI curricula into undergraduate medical education, including licencing requirements, limited curricular hours, and a lack of faculty expertise in AI. However, a standardised framework of competencies for implementing a unified AI curriculum is not yet available.

The Faculty of Medicine at Ulm University, like many other educational institutions, has a long-standing expertise in e-learning and digital skills. The application of these skills has undergone a significant and positive evolution due to the COVID-19 pandemic [46]. Recently, a questionnaire was administered to gather more data on university teacher´s digital literacy [47]. However, the integration of AI-tools into medical education remains a relatively new topic, necessitating detailed information about students' current AI knowledge across various programs. This information is essential for developing and implementing a suitable AI curriculum that ensures continuous medical education in this emerging field. To gain insights into how medical students are using AI tools, both privately and in their studies, a cross-sectional online questionnaire was conducted. This study aimed to assess their *knowledge, skills, and attitudes* related to AI tools like ChatGPT, thereby describing the current status at the Faculty of Medicine at Ulm University. Additionally, the data will address the following research questions:

(1) Are there any differences in responses between medical students (state examination program) and students studying in life science courses (bachelor or master program)?
(2) Can differences be found between German-speaking students and those with another native language?
(3) Are there any differences in responses between male and female students?
(4) Is there a correlation between response behaviour and students' subjectively assessed performance level or AI expertise?

Material and Methods

Study Design

A monocentric questionnaire-based cross-sectional study was conducted in order to assess students' knowledge, prior experiences (skills), and attitudes at the Faculty of Medicine in Ulm. The survey was carried out in a voluntary and anonymous manner to all registered medical and dental students, as well as those enrolled in the international Bachelor's (BA) and Master's (MSc) degree programs in Molecular Medicine (MM) and Translational Neurosciences (TN). A study proposal was submitted, and the university's ethics committee granted a formal waiver (decision date: July 12th, 2023). Students received additional information about the survey's content through the Unipark platform (Enterprise Feedback Management Software, Tivian XI GmbH, 2021, Köln), and their written consent was obtained prior to responding to any questionnaire items. No incentive was offered or provided to participants in the survey.

Students at the Faculty of Medicine in Ulm were surveyed online during the ongoing semester (November 2023) following an email invitation from the Office of Students' Affairs. This email, included an access link to the Unipark survey platform, which was available for a period of four weeks. Each student could participate only once using an internet-enabled device (either mobile or PC). Additional invitations were disseminated through projections in the main lecture halls and on campus media screens, and email reminders were sent twice at regular intervals throughout the four-week period.

There were 3995 officially registered students at the Faculty of Medicine [48]; however, not all students in the specified degree programs could be invited due to organisational reasons (for example, not every student had re-registered for the winter term). Table 1 presents the number of registered students at the Faculty of Medicine who were invited to participate, along with the number of respondents and overall response rates after applying the pre-defined inclusion criteria (verified consent, only single missing values, and log files showing handling times of the questionnaire of less than 5 minutes).

Table 1. Number of students at the Faculty of Medicine and study participants categorised by degree program, along with the resulting survey return rate (rounded values).

Number of students Faculty of Medicine						
	Total	Sex (m/f/)	Nationality (German/Abroad)	Invited students	Participants (N/%)*	Return rate
Medical students	2997	1379/2616	3478/517	2816	170/82.5%	6.0%
Dental students	436	146/290	387/49	357	9/4.4%	2.5%
Molecular Medicine	403	83/320	304/99	229	22/10.6%	9.6%
Molecular and translational neuroscience	53	12/41	30/23	58	5/2.4%	8.6%
Sum	3995	3995	3995	3460	206/100%	**6.0%**

*After exclusion of 3 participants; 1 student did not indicate the degree program.

Questionnaire Development and Design

The questionnaire's content was informed by an initial bibliographic search conducted in May 2023, using keywords such as "Artificial intelligence and medical education", "AI in medicine", "AI and medical students", "AI readiness", "ChatGPT and healthcare students", as well as the scientific publications cited by Sallam in his systematic review "ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns" [6]. The search aimed to create a questionnaire lasting 10 to 15 minutes, assessing students' current knowledge and competence related to AI and ChatGPT. Two authors (A.B., L.M.) then collected, categorised, and evaluated individual survey items using the SPSS method recommended by Helfferich [49].

Since a specific ChatGPT literacy scale was not available during the development of the questionnaire, the authors referenced the German version of the by Karaca et al. [42] developed "Medical artificial intelligence readiness scale for medical students" (MAIRS-MS) which was translated by Laupichler et al. (2022) [43].

Four out of 22 items –question 9,12,18 and 21- were adopted, covering the main topics of "cognition, ability, vision, and ethics". The study and questionnaire design were discussed in collegial coaching sessions with university colleagues who have extensive expertise in questionnaire development (A.S) and research in medical education and digital learning (J.K.). Finally, the questionnaire was integrated into the Unipark online platform and tested by nine students from the aforementioned degree programs during a one-week pilot project to evaluate comprehension, spelling, and technical functionality.

The final questionnaire encompassed a total of 33 individual items or item groups related to the MAIRS-MS, ensuring both quantitative and qualitative data collection. Along with demographic data ($n = 6$), respondents were asked to evaluate their previous knowledge ("cognition") ($n = 5$) regarding AI, describe their practical experience with AI tools ("skills") ($n = 5$), and specifically with ChatGPT ($n = 4$). They were also asked to assess the risks or weaknesses ($n = 3$) and opportunities or benefits ($n = 2$) of AI. Additionally, four items addressed students' attitudes and opinions about AI. Finally, students were asked about their desire for further educational support and their beliefs regarding potential future applications of AI ($n = 4$). The questionnaire was available in both German and English. The English version is included in the appendix (**Supplement 1**). On average, participants took 13 minutes to complete the online questionnaire.

Responses were quantified using a five-point Likert scale: 1 (does not apply at all) to 5 (fully applies). In some cases, multiple answer options were permitted. Open-ended responses, such as "other AI tools used in studies", "other advantages and opportunities" or "other tasks" were transcribed, categorised and quantified afterwards.

Statistics

The Unipark online software directly transferred all survey data into a SPSS spreadsheet. Data were analysed using SPSS Statistics software (IBM Corp. Released 2022. IBM SPSS Statistics for macOS, Version 29.0. Armonk, NY: IBM Corp). Quantitative data are presented as median values, and for better comparability, also as mean ± standard deviation (MD; MN ± SD). In some cases, minimal (MIN) and maximal (MAX) values were included. Internal consistency was expressed using Cronbach's α. All data were analysed for distribution patterns using the Kolmogorov-Smirnoff test. For subgroup analysis, ordinal scaled data (e.g. Likert scaled) or non-Gaussian distributed data were analysed using non-parametric tests: Mann–Whitney U-test or the Kruskal–Wallis–H test (for more than two groups), with post-hoc comparison by Dunn's multiple-comparison test. The level of statistical significance was set at 5% ($p = 0.05$*, $p < 0.01$ **, $p < 0.001$***).

Effect sizes were expressed by Cohen's *d*, with values of 0.2, 0.5, and 0.8 indicating small, medium, and strong effects, respectively. Subgroup analysis for nominal scaled data were performed using Pearson-Chi Square Test or Fisher's Exact Test if the sample size was smaller than $n = 30$. The effect size for Chi-Square Test's effect size was calculated using Cramer's V, with values of 0.1 0.3 0.5 indicating small, medium, and strong effects. The Spearman coefficient was used as a correlation measure for ordinally scaled data, with $|\rho| = 0.10$ considered weak, $|\rho| = 0.30$ moderate and $|\rho| = 0.50$ strong.

Transparency Statement

This manuscript has been optimised in English wording and style using the AI tools "DeepL" and "ChatGPT", and no other AI tools were applied during the preparation of the manuscript.

Results

Of the 3460 students contacted by the Faculty of Medicine 6% ($n = 207$; 62.4% female; 36.6% male; 1% unspecified) participated in the study. Three questionnaires were excluded from the data analysis due to non-compliance with the inclusion criteria. Internal consistency using Cronbach's α for questions applied form the MAIRS-MS questionnaire showed values of $\alpha = 0.782$, for the subscale knowledge ($n = 24$ items) of 0.829, for the subscale skills ($n = 15$ items) of 0.917 and for the subscale attitudes ($n = 12$ items) 0.745.

The participating students had an average age of 23.2 years (SD ±4.5; MIN 18; MAX 51). The vast majority were enrolled in a medical degree program (82.2%) and were, on average, in their fifth semester (MIN 1st MAX 15th). A small percentage (8.3%) reported that they were not native German speakers. When asked to self-assess their overall academic performance, the majority classified themselves as average students (66.7%), while 14.5% considered themselves below average and 18.8% rated themselves above average. Over half of the participants (53.6%) indicated that they had previously authored some form of scientific paper independently. Regarding their current understanding of "artificial intelligence" the majority rated their knowledge as "sufficient" (MD 4 = sufficient; MIN 1 = very good; MAX 6 = unsatisfactory).

When asked about the fundamental technology underlying AI, a quite large proportion of students acknowledge that they had either never heard of it or were only somewhat familiar with it (63.3–91.3%) (Figure 1).

About two-thirds of the students reported familiarity and prior use of ChatGPT (66.7%). This was followed by AI tools like DeepL Translater (42.0%), Quizlet (35.7%) and Grammarly (17.4%). Most other listed AI tools – including ChatPDF, Conker, Connected Papers, DALL-E, DeepL Write, ProWritingAid, Elicit, Peplexity, PesearchRabbit, Midjourney, StableDiffusion, and SciSpace – were unfamiliar to over 85% of respondents. Tools specific to the medical context such as ADA (3.3%) and

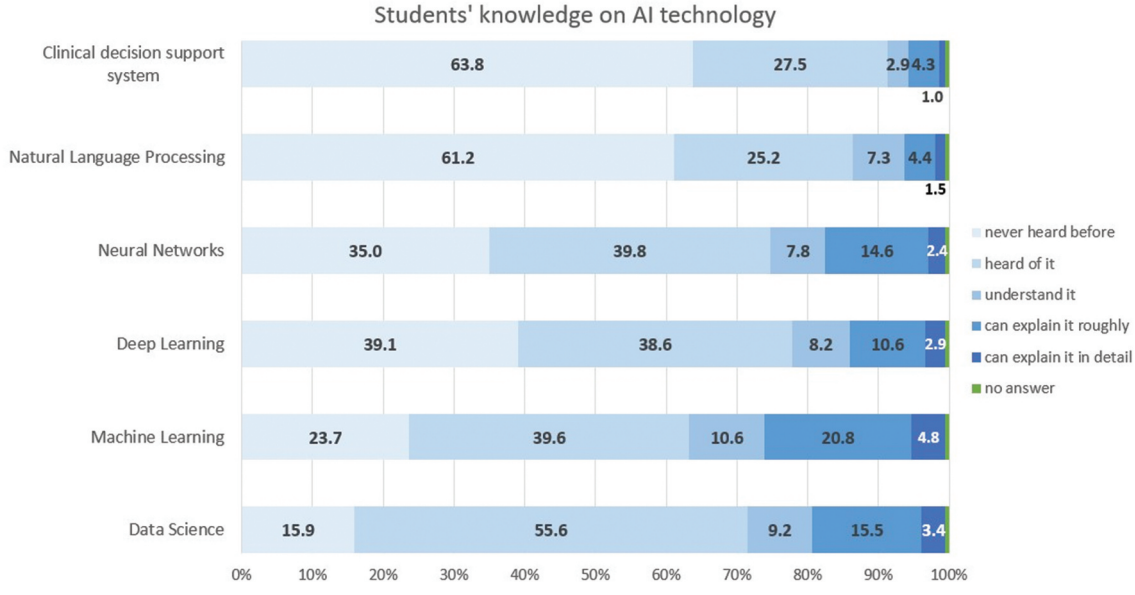

Figure 1. Students' responses ($n = 207$) to the item "please assess your current knowledge on the following AI topics" using a 5-point likert scale plus an option not to answer at all.

DocMedicus (15.4%) were recognised by only a minority. Students also mentioned additional AI tools in open-ended responses, including "*BING-AI*", audio applications like "*Fireflies*", "*Auphonic*", or game apps for chess. However, the frequency of use for these tools were – aside from ChatGPT – relatively low, with only half of the students using them at least once a month.

On a 5-point Likert scale (1 = does not apply at all, 5 = completely applies), respondents rated their skills in using AI tools at an ambiguous average. Students recognised AI's strengths in accessing knowledge (MD 3.5; 3.31 ± 1.22) but felt relatively weak in creating new knowledge using these tools (MD 2; 2.3 ± 1.22). Similarly, one-third of the students either disagreed or were undecided about explaining the pros and cons of AI technology (MD = 3).

A more detailed question asked students to quantify their use of AI tools for specific tasks. Unsurprisingly, only a minority of medical students used AI tools regularly for data analysis (10.3%) or programming (5.8%). However, a significant number applied AI tools as simple search engines (43.2%), for explanations (39%), translations (25.9%), discussing problems (18.9%), or literature searches (18%). A surprising low number of participants reported using AI tools like ChatGPT for analysing, editing or generating text. Qualitative responses ($n = 3$) indicated that students used ChatGPT to create MCQs or flashcards (Figure 2).

Only 17.9% ($n = 140$) of those participants familiar with ChatGPT had consulted guidelines for effective prompt creation. A significant portion (40%) developed their prompting approach through trial and error, while nearly a third had not considered prompt settings at all, and 11.4% lacked experience in systematic prompt editing.

Regarding the quality of ChatGPT-generated texts, participants mostly evaluated them accurately. They recognised the need to critically review ChatGPT texts (MD = 4; 4.41 ± 0.73), and its cited references (MD = 4; 4.3 ± 0.83), noting that ChatGPT often produces inaccuracies (MD = 3; 2.82 ± 0.91). They also acknowledged ChatGPT's ability to generate high-quality conversational examples (MD = 3; 3.22 ± 0.86) with few spelling or grammatical errors (MD = 2; 2.59 ± 0.93). However, there was considerable uncertainty about the legal framework governing AI tools use for academic writing. More than half of the respondents felt unable to assess the accuracy of the questionnaire statements, yet 61% recognised that they could not use AI tools without restrictions (Table 2).

Furthermore, students largely agreed that AI would not render medical staff dispensable (MD = 2; 1.94 ± 1.15) but acknowledged potential data bias (MD = 4; 4.05 ± 0.92). They also identified AI weaknesses, including inadequate data security, unclear liability issues, and insufficient representation of real-life medical

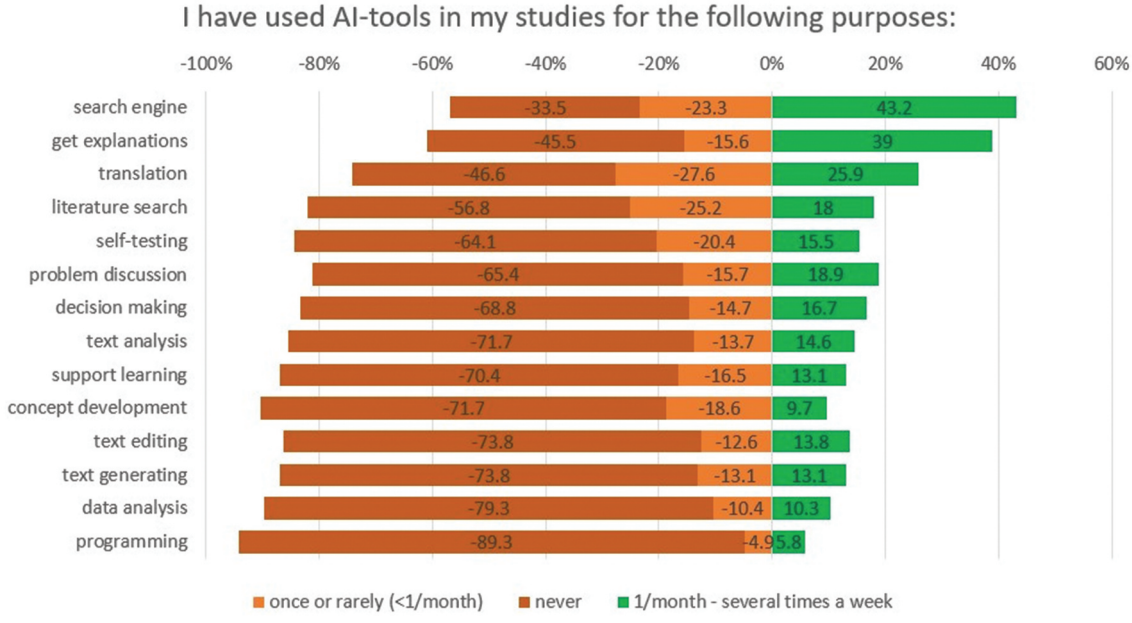

Figure 2. Students' responses ($N = 207$) to the item "I have used AI tools in my studies for the following purposes: [. . .]" presented on a 7-point scale (1 = never, 2 = once, 3 = rarely, 4 = once a month, 5 = several times a month, 6 = once a week, 7 = several times a week). A frequency of less than once a month (point 1–3) was classified as non-users (red/orange color) versus students that used AI-tools more than once a month (point 4–7) (green color).

Table 2. Percental distribution of students' assessment regarding statements related to the legal framework of AI tool application within the university context. Based upon the author's experience and local university regulations, all statements except the first one are considered to be mainly correct.

Item	Correct	Incorrect	Don't know
I may use ChatGPT/AI tools for writing a housework, term papers or my thesis without any restrictions.	3.4%	61.0%	35.6%
I may use ChatGPT/AI tools for writing housework/term papers/thesis only in accordance with current study regulations.	48.3%	10.7%	41.0%
The University may ask me to submit a self-declaration that confirms that the main work of my term paper/thesis was done without ChatGPT/AI tools.	43.9%	7.3%	48.8%
Lecturers may ask me to provide a list of all AI tools used in a housework/term paper/thesis.	40.5%	6.3%	53.2%
I have to name ChatGPT as an author in the bibliography of my housework/term paper/thesis or scientific paper.	15.6%	19%	65.4%
By marking text elements with " … " I may quote ChatGPT content verbatim, just like citing from any publication.	7.8%	27.8%	64.4%

*n=205, 2 missing values.

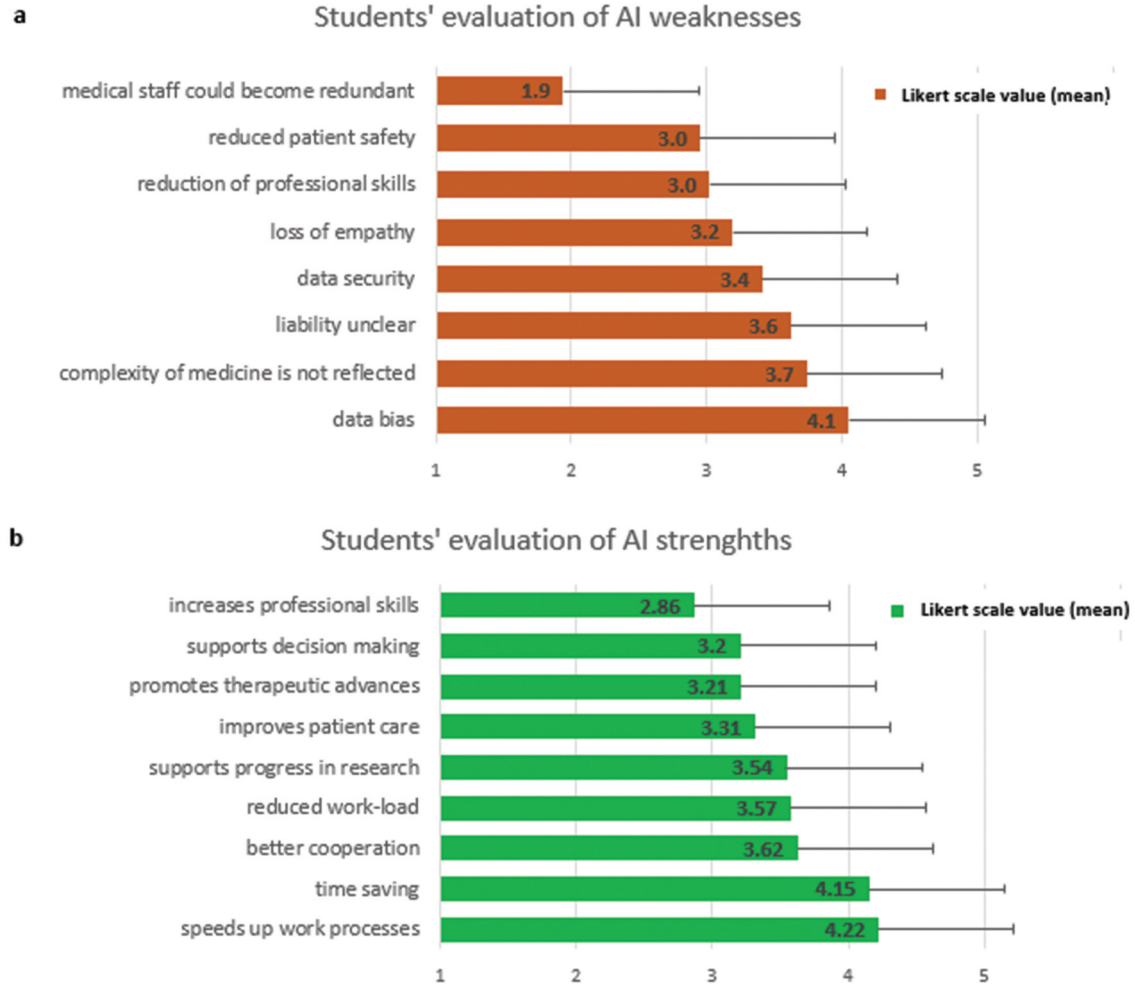

Figure 3. Students´ responses (n = 207) to the item "In which aspects do you see the greatest risks or weaknesses of AI?" (Figure 3a) and "Which of the following aspects do you consider as the greatest opportunities or advantages of AI" (Figure 3b). Both items were evaluated on a 5-point likert scale (1 = does not apply at all, 5 = fully applies) plus an additional option not to answer at all. Results are given as mean value and standard deviation.

complexities (Figure 3a). Students recognised AI's benefits in accelerating workflows (MD = 4; 4.22 ± 0.916), saving time (MD = 4; 4.15 ± 0.927), reducing workload (MD = 4; 3.57 ± 1.1), and enhancing collaboration among clinics, physicians, and scientists (MD = 4; 3.62 ± 1.08). However, there was considerable uncertainty regarding AI's role in facilitating faster research progress, leading some students (n = 22) to opt out of answering these questions (Figure 3b). Qualitative responses (n = 4) highlighted concerns like "*unverified false diagnoses*", "*money invested would be better spent on empathic, human medicine*", "*lack of interpersonal exchange in cases of mental*

illness", "*AI as a basic instrument, physicians have to check its results*".

Participants were asked to evaluate their ability to use AI tools in accordance with ethical principles (an explanatory short note on "ethical principles" was provided in the questionnaire). The median score on the 5-point scale was 4 (largely applies) (MN 3.63 ± 0.868).

The questionnaire revealed that students generally hold positive attitudes towards the application of AI tools in medicine. They expressed more agreement with feeling of excitement, openness, curiosity, and interest, while showing less agreement with feelings of panic, fear or hesitation (**Supplement 2:** Spider chart representing students" responses to the item "My current view on the topic "AI tools in medicine" can be described by [...].). Some free-text responses ($n = 5$) revealed additional emotions, including confidence, anger, concerns, mistrust and scepticism.

The majority of respondents held a positive attitude towards implementation of AI in medicine (MD = 3; 3.39 ± 0.96) and believed its future use is essential (MD = 4; 3.53 ± 1.07). However, only 41% strongly agreed (Likert scale level 1 and 2) that AI will revolutionise their future career (MD = 3; 3.27 ± 1.05), and the same percentage felt that acquiring basic AI skills is crucial for competitiveness (MD = 3; 3.19 ± 1.09). Notably, over half of the participants (62.2%) disagreed with the statement that they feel confident using AI tools professionally (MD = 2; 2.32 ± 1.16).

Overall, 63.2% of students expressed a desire to further their education in AI related to medicine, while 27.5% showed clear interest and 9.3% no interest in additional training. Some free-text responses indicated practical priorities, such as *"I am taking my state examination first"* or "*I am taking modules at university computer science courses'*.

Regarding self-directed learning, students reported acquiring digital competencies primarily through trial and error, online platforms, or help from classmates (Figure 4). Only about a quarter mentioned using publications, conferences, or professional training opportunity.

Less than 10% of respondents showed no interest in further training on "AI in general". Students expressed strong interest in several AI-related topics: appropriate AI tools for their studies (95.5%), academic writing (91.9%), future medical applications (94.5%), legal aspects of AI (94.4%), and interpreting AI in (clinical) decision-making (90.5%). However, about a third were not interested in further training on integrating AI in teaching contexts, such as exams or classroom instructions (Figure 5).

The most frequently mentioned topics were analysed to identify what kind of training opportunities students desire. Overall, the number of positive responses decreased considerably from general requests for information (e.g. newsletters, information events) with 1006 nominations, to voluntary formats like elective courses or online tutorials (946 nominations) and finally to mandatory curricular courses (362 nominations). The top three AI topics students wanted information on were legal aspects on AI application (58.1%), current research and development in AI (55.5%), and AI tools for scientific writing (55.1%). For voluntary learning opportunities, students showed strong interest in using AI tools in their studies (62.5%), AI in general (58%) and AI in scientific writing. Although mandatory

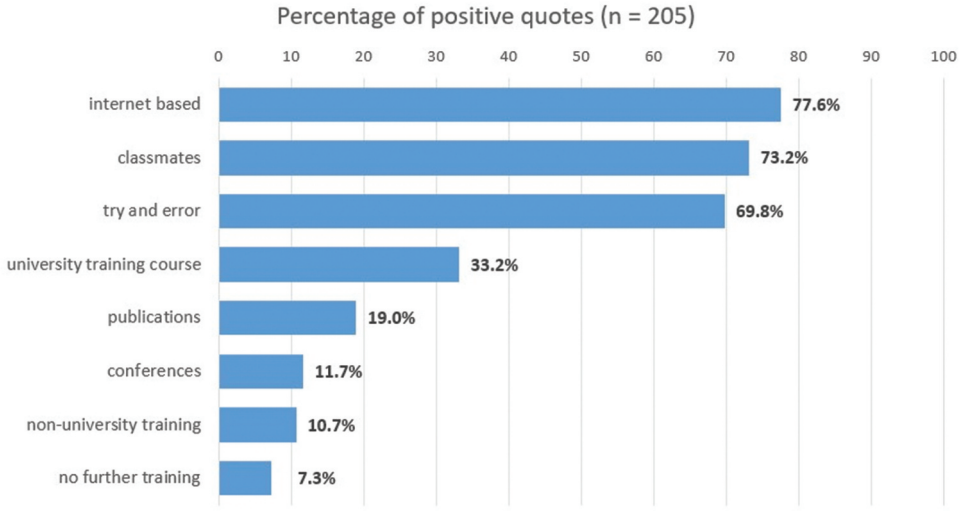

Figure 4. Percentage of positive quotes of $n = 205$ students (100%) related to each answer option to the item "How do you acquire new knowledge and skills in the field of digital competence?". The item included a multiple answer option revealing in summary $n = 620$ quotes. Data are displayed as percentages.

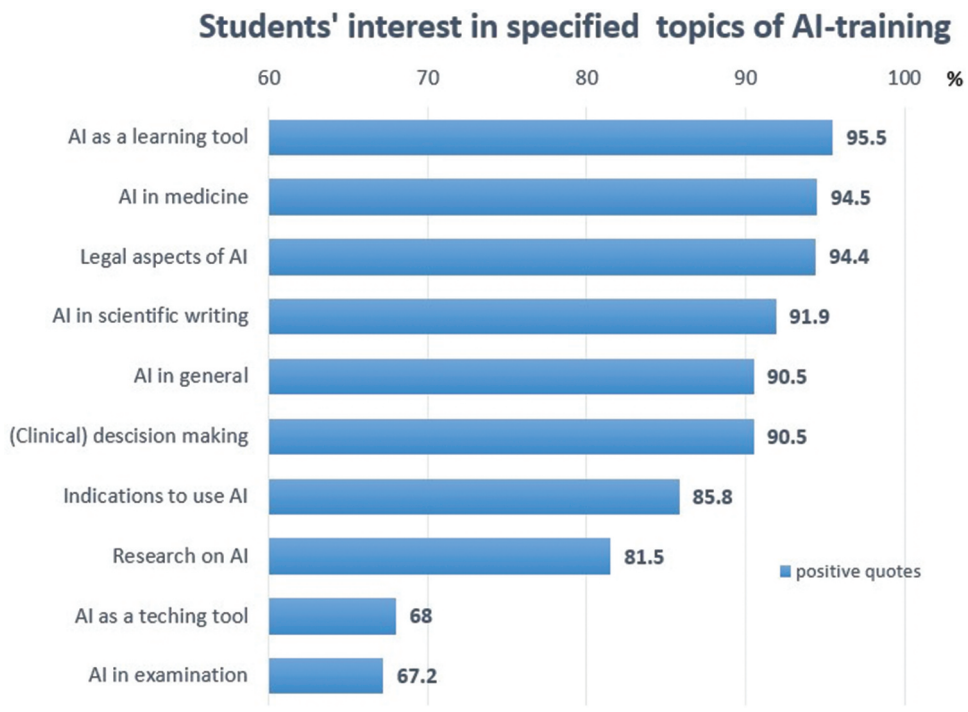

Figure 5. Responses of students (n = 207) to the item "On which (of the following) topics would you like to receive information or support (…)?". Data are displayed as percentages of positive quotes.

training courses were less favoured, the most frequently requested topics for required education included: evaluating and interpreting AI output in decision-making (32%), legal aspects related to AI (30.8%) and future uses of AI in medicine (23.4%) (Table 3).

Subgroup (SG) Analysis

Is there a significant data difference between medical students and students studying other life science courses?

Most of survey participants were medical or dental students at Ulm University (SG 1: 87%, $n = 180$), while the remaining 27 students (SG 2: 13%) were in Bachelor's or Master's programs, including Molecular Medicine and Translational Neurosciences. Few significant subgroup differences emerged. Life science students (SG 2) reported using DALL-E* tool more often (14.8% vs. SG 1: 3.3%; $p = 0.038$, Cramers V (V) = 0.186) and were less likely to report never using AI tools for text editing (SG 2: 55.6% vs. SG 1: 76.1%; $p = 0.006$; $V = 0.333$). No differences were found between the groups in AI-related knowledge, experience in writing scientific writing, or frequency of AI tool and ChatGPT usage. Both groups held similar views on handling AI-generated texts and showed comparable preferences for training formats on AI use.

*DALL-E, https://openai.com/index/dall-e-3/ Imagegenerator, OpenAI, 2021

Is there a sigsnificant data difference between German speaking students and students with another native language?

Table 3. List of students' quotes about their interest in further education on specific AI topics and the preferred teaching format (multiple answer option). The positive responses of the students registered in the Faculty of Medicine in ulm (n = 207) are listed. The top three AI-topics with highest percental votes in each column are marked by bold printing.

	AI topic	No interest in training*	Information e.g., newsletter	Voluntary e.g., elective	Mandatory e.g., seminar
1	AI as a learning tool	9	106	**125**	32
2	AI in medicine	11	103	107	**47**
3	Legal aspects of AI	11	**115**	76	**61**
4	AI for scientific writing	16	**109**	**114**	33
5	AI in general	19	102	**116**	33
6	AI (clinical) decision making	11	92	97	**64**
7	Indications to use AI	28	100	94	29
8	Research on AI	37	**111**	81	19
9	AI as a teaching tool	64	82	73	20
10	AI in examinations	66	86	63	19
	Number of positive responses	**280**	**1006**	**946**	**362**

Among 207 respondents, only 17 (8.3%) identified as non-native speakers (SGnn), two of whom were in the Translational Neuroscience program (nationwide foreigner student quota: 5%). Some subgroup differences were identified: non-native speakers were more likely to be familiar with Grammarly* (35.3% vs. 15.9%; $p = 0.029$; $V = 0.187$) and more likely to report a more frequent use of DeepLTranslator** ($p = 0.013$; $V = 0.444$).

*Grammarly, https://grammarly.com Grammarly Inc., 2009, San Francisco, USA.

**DeepL SE, https://www.deepl.com Online-Translating Service, 2017, Cologne, Germany.

Non-native speakers rated their AI proficiency slightly higher, including more frequent use of AI tools for online searches ($p = 0.049$; $V = 0.286$). Concerning their prompting skills, native speakers tended to enter prompts impulsively, whereas non-native speakers reported refining their approach through trial and error (MD = 3 vs. SGnn: MD = 2; $p = 0.02$; $V = 0.241$). The two groups had similar views on AI's pros and cons. Non-native speakers also expressed more confidence in using AI professionally (MD = 3 vs. SGnn MD = 2; $p = <0.001$; $V = 0.386$) and a more positive attitude (MD = 4 vs. Snn MD = 3; $p = 0.045$; $V = 0.218$), showing greater enthusiasm (MD = 4 vs. SGnn MD = 3; $p = 0.01$; $V = 0.256$) and less hesitation (MD = 2 vs. SGnn MD = 3; $p = 0.034$; $V = 0.225$). Interest in further AI training was similar across both groups.

Is there a significant data difference between male and female students?

Of all survey participants, 126 identified as female (62.4%), with 60% of them having used ChatGPT. Among the 74 male participants, significantly more (81.9%) reported ChatGPT use ($p = 0.002$; $V = 0.228$). Female participants rated their general AI knowledge lower than male participants ((6 = very good; 1 = unsatisfactory) f: 4.33 ± 1.12; m: 3.45 ± 1.31; $p = 0.016$, $V = 0.262$), though both genders rated their academic performance similarly.

Differences also emerged in AI tool usage: female participants reported lower frequency of AI tool use across various tasks. For instance, 58.1% of male participants had never used AI tools for text analysis, compared to 77.8% of females ($p = 0.041$; $V = 0246$). While both genders expressed emotions like panic, openness or curiosity towards AI, female participants reported fear ($p = 0.008$; $\rho = -0.197$**, Cohen $d = 0.411$), hesitation ($p = 0.011$; $\rho = -0.175$*) more often, but were less likely to report interest ($p = 0.010$, $\rho = 0.186$**) or enthusiasm ($p = 0.024$, $\rho = 0.148$*) **(Supplement 2)**. Both groups shared a general interest in further AI training. Regarding digital competence, female students were less inclined to learn through trial and error (62.1% vs. 81.1%; $p = 0.017$, $V = 0.202$), preferring instead to seek help from peers (79.8% vs. 60.8%). No significant gender differences emerged in preferred training topics or format, though male students were more likely to favour mandatory AI training (25.4% vs. 11.5%; $p = 0.035$, $V = 0.185$) and AI tool use in academic writing (28.2% vs.10.7%; $p = 0.007$, $V = 0.225$).

Is there a significant data difference correlating with the subjectively assessed level of performance or AI expertise?

Level of performance:

Students categorized themselves by self-estimated academic performance into three groups: weak (SG 1:14.5%), average (SG 2: 66.7%) or strong performer (SG 3:18.8%). Significant differences were noted in AI knowledge, with weaker students rating their prior AI knowledge lower than stronger students ((6 = very good; 1 = unsatisfactory) SG 1: MD = 5; 4.5 ± 1.11 vs. SG 3: MD = 4; 3.54 ± 1.3 $p = 0.005$, Cohen's $d = -0.786$). While strong performers generally reported more familiarity and prior use of AI tools, but significant subgroups differences appeared only for DALL-E * ($p = 0.001$), DeepL Write** ($p = 0.027$), and Quizlet*** ($p = 0.047$), not for ChatGPT. Stronger performers were also more likely to use AI tools more than once per month, particularly ChatGPT ($p = 0.006$, $\rho = 0.267$**), DeepLWrite ($p = 0.015$, $\rho = 0.541$), and Grammarly ($p = 0.010$, $\rho = 0.240$).

**DeepL Write, https://www.deepl.com/de/write Writing Assistant, 2017, Cologne, Germany.

*** Quizlet, Quizlet Live https://quizlet.com/de/features/live, Quizlet Inc. 2007, San Francisco, USA.

Strong performer (SG 3) rated themselves as more capable of accessing (SG 1: MD = 3; 3.07 ± 1.22 vs. SG 3: MD = 4; 3.79 ± 1.28; $p = 0.015$) and sharing factual data using AI tools (SG 1: MD = 3; 2.9 ± 1.21 vs. SG 3: 3;3.36 ± 1.181; $p = 0.033$). Across all groups, AI tools were most often used as a search engines, with usage rates of 46.7% for SG 1, 70.3% for SG 2, and 69.2% for SG 3, while programming applications were less common (SG 1 3.3%, SG 2 10.9%, SG 3 20.5%). Lower-performing students reported less frequent use of AI tools, and 9 of the 14 tasks showed statistically significant differences among subgroups, with effect sizes ranging from 0.42 and 0.91 (**Supplement 3:**Subgroup analysis of low, medium and high performing students according to their self-estimation concerning the frequency of AI tool application.)

The subgroups showed similar self-assessments in prompt engineering and awareness of AI risks and opportunities. However, lower-performing students

(SG 1 and SG 2) were more likely than SG 3 to believe that ChatGPT-generated texts often require corrections for spelling or wording errors ($p = 0.003$). Both SG 1 and 3 rated themselves as more able to use AI in an ethical manner (MD = 4, "largely applies") compared to average performers in SG 2 (MD = 3, "partially applies"; $p = 0.024$). Students with lower self-reported performance felt less confident in applying their AI skills professionally than those in SG 3 (1.97 ± 1.15 vs. 2.63 ± 1.17; $p < 0.05$; $d = 0.568$). In contrast, SG 3 students had a more favourable view on AI's role in medicine (MD 4 vs. 3; $p = 0.009$). No significant subgroup differences emerged regarding attitudes towards AI's implementation in medicine, its impact on future careers, or interest in further AI training.

AI expertise:

Students rated their AI knowledge from 1 (very good) to 6 (unsatisfactory), forming three groups: lower self-assessed AI knowledge (SG-low, grades 5–6, 13%), average (SG-middle, grades 3–4, 44%), and higher (SG-high, grade 1 and 2, 43%). This classification aligned with significantly differences in subgroup knowledge of AI-related topics (e.g. data science), as well as their self-rated skills with AI tools ($p < 0.001$, correlations ranging from 0.276** to 0.508**). AI expertise and academic performance were also moderately correlated ($\rho = 0.236$**), though no link was found between AI knowledge and age.

SG-high students reported more frequent and confident use of AI tools to access and evaluate knowledge (MD 4 vs. 3) and could better explain AI's strengths and limitations (MD 4 vs. MD 2 for SG-low; $p < 0.001$). AI tools were used significantly less by SG-low students than by SG-high ($p < 0.001$), except in cases where AI supported learning or self-assessment. While 92.3% of SG-high students had used ChatGPT, only 51.1% of SG-low had done so ($p < 0.001$).

Prompting methods also varied significantly across subgroups ($p < 0.001$) with SG-low students generally lacked a systematical prompting approach (MD = 4 vs. SG-high MD = 2). SG-high students perceived fewer risks in data security ($p < 0.001$) and dehumanisation ($p = 0.004$), but saw AI's opportunities similarly to other subgroups.

SG-high students, compared to SG-low, rated AI as more important for their future careers (MD = 4 vs. 3) and felt more confident (MD = 4 vs. 2) and positive about AI in medicine (MD = 4 vs. 3; $p < 0.001$). Although levels of panic and fear were similar, SG-high students reported more excitement, openness, curiosity and interest. Subgroups differed in preferred learning methods for digital competence: 63% of SG-high intended to further their education in AI, while only 22.7% of SG-low expressed this intention ($p < 0.001$, $V = 0.222$). SG-high students preferred trial-and-error learning (88.9% vs. 54.4% in SG-low; $p < 0.001$; $V = 0.295$) and were more likely to attend scientific meetings (25.9% vs. 5.7%; $p = 0.015$, $V = 0.205$). Conversely, 84.1% of SG-low students relied on peer support, compared to 48.1% in SG-high ($p < 0.001$, $V = 0.265$).

Training topics and format preferences were similar across subgroups, though SG-low showed a higher disapproval rate for mandatory AI courses than SG-high (82.6% vs. 65.4%; $p = 0.013$). Similar trends were noted for topics like research and development of AI, future use of AI in medical applications, AI integration as a teaching tool, and AI usage guidelines.

Discussion

Main Results

This study analysed survey data from students at a German medical faculty, assessing their self-rated knowledge, usage pattern, and attitudes. While most students were familiar with ChatGPT, other AI-tools were significantly less known or used. Students generally took a pragmatic approach, using AI tools primarily as search engines or for simple information queries rather than for medical or scientific applications. Despite moderate knowledge of AI, many expressed uncertainties in practical AI skills such as prompt engineering and understanding its legal framework. Although respondents were open to integrating AI into their studies and careers, they felt underqualified to use AI professionally and were therefore interested in further training – ideally within voluntary, study-relevant formats.

According to the previously defined research questions subgroup differences emerged across key areas:

a. **Non-native speakers** showed a more positively attitude towards AI-tools in medicine. They had used AI tools more frequently for text translation or editing and reported more experience with systematic prompting techniques than native speakers.
b. **Female students** had used AI tools less frequently and reported feelings of fear or hesitation towards AI implementation more often than male students.
c. **Lower-performing students** felt less confident in their AI skills and showed more negative attitudes towards AI compared to high-performing students, who used AI tools more frequently and for more complex tasks.

d. **Students with less AI knowledge** were more likely to misjudge AI-generated text quality and held more negative views on AI.

Overall, subgroups showed similar interest in further AI training, indicating a broad need for accessible, relevant educational resources in AI across performance and demographic lines.

Limitations

This study's findings, based on single-point data from one institution, may not fully be generalised to other settings. Given AI's rapid evaluation, these results represent a snapshot in time. Only 6% of invited students participated, introducing a potential selection bias and therefore might not represent the full student cohort; by comparison typical online survey response rates hover around 44.1% [50]. However, similar to our response rate Stewart et al., 2023, published survey data on medical students' attitudes towards AI after gaining a response rate of 8.9% [51]. Incentives could have increased participation but were intentionally avoided to minimise a potential self-selection bias.

By majority medical students participated in this survey, with only a limited representation of students attending Bachelor's or Master's programs, suggesting that subgroup findings should be viewed cautiously. Furthermore, the uneven semester distribution may also affect the accuracy of students' self-reported knowledge, skills and attitudes, as age and academic stage were not fully controlled for in the study design. Finally, the questionnaire, though reliable, was self-developed and had not been validated in advance. However, its development followed solid guidelines and ultimately showed good internal consistency. Data gained in this questionnaire capture only students' self-estimated AI knowledge, skills, attitudes or their self-estimated study performance rather than their actual competence or performance.

Discussion of study results with Reference to the Current Literature

The findings from this study align with prior research on students' AI knowledge, skills and attitudes, which consistently show considerable uncertainty in handling AI and students' limited understanding of foundational AI principles. Similar trends were noted in studies from Turkey [45], Spain [52], Australia ,[51] and a multi-centre international study [53]. Pinto dos Santos et al. (2019) found that 68% of German medical students were unaware of AI use in radiological diagnostics [54]. However, this study was conducted before ChatGPT's release, and ever since general AI awareness has risen. Recent German studies published post-ChatGPT"s release support these findings. Von Garrel et al. (2023) surveyed 6300 German students across various disciplines and reported that 63.4% already use AI tools in their studies [55]. With usage highest among engineering students (75.3%) and lowest among health science/medical students (52.7%). ChatGPT was used by 50% of students across programs, with AI tools primarily applied to clarify questions and to explain discipline-specific concepts.

The data from this study of medical students at Ulm University can be compared with a university-wide survey conducted at the same time (autumn 2023), which assessed student's knowledge and skills regarding AI tools across all degree programs [56]. This survey included 2136 students with a response rate of 29%, of whom 613 were medical students. While approximately two-thirds of participants in von Garrel's study reported using AI tools, only about a half of Ulm's students had done so. Furthermore, 47% indicated lacking knowledge on AI tools.

The results shown here indicate that medical students are familiar with only a few AI tools, primarily ChatGPT, which they use only infrequently for text editing or creation. Instead, they utilise it mainly for tasks of lower taxonomy such as translation, searching for information or clarifying questions. Consequently, students lack the skills needed to fully leverage the potential of LLMs like ChatGPT in professional settings, as envisioned by experts in research [57], teaching [58] and in the context of patient management or clinical decision support [57,59,60]. This "misuse" of AI may stem from students' underdeveloped skills in systematic prompting, highlighting the need for enhanced training programs in both, university and school curricula.

Before discussing students' preferences for further training, it is important to consider this study's subgroup analyses: The findings revealed gender-specific differences in students' attitudes towards AI integration in medicine. Notably, females exhibited more negative attitudes and greater feelings of fear or hesitation regarding AI. This aligns with prior study results for students from Germany [54] and Spain [52]. The negative attitudes among females may be linked with their lower usage of AI tools leading to decreased familiarity and confidence in these technologies. This study found no significant differences in AI tool usage among medical, dental or life sciences bachelor's and master's students. This may be due to the small number of life sciences participants, because von Garrel's larger

survey identified a correlation between usage behaviour and degree program, indicating that master and doctoral students utilise AI tools more frequently for writing scientific or academic papers; a practice not mandatory in medical or dental programs in Germany.

In this study, nearly 10% of students expressed no interest in further AI training. This supports prior findings, where 91% of medical students favoured AI training, and most students supported its integration into their curriculum [51,54]. Training in AI is essential for the future [52]. To complement previous research, this survey also explored students' preferred content and teaching formats for AI training: Students were particularly interested in topics that could be directly applied to their daily tasks, such as using AI to enhance their learning, assist in patient care, and aid in scientific writing. They also sought on the legal use of AI tools [56]. However, respondents generally rejected the idea of AI being taught as a compulsory subject. Overall, while students are eager to harness the benefits of AI, they express significant uncertainty about its professional application. Despite their interest, their willingness to engage in compulsory or time-consuming training formats appears limited.

Interest in AI extends beyond students; educators and healthcare professionals have already highlighted the need to update medical curricula to incorporate AI [31,38,45,61]. Yet, medical educators' digital competencies are often described as inconsistent and moderate to low. Like students, educators mainly acquire digital skills independently from peers [47]. Digital skills training is largely absent from the mandatory core curriculum and only recommended in Germany's optional learning objectives framework, NKLM 2.0 [62]. Therefore, educators' motivation for additional training has been rather modest upto now, pointing to a broader need for improved digital qualifications.

Despite AI's growing use in clinical settings, comprehensive AI education remains lacking in German medical training. A 2022 survey by Mosch et al. reported that 71.8% of the German medical faculties include AI in their curricula, though typically only as elective courses – an approach also preferred by the students in this study [63]. Currently, AI is sporadically covered in undergraduate courses [23,58], and none of Germany's 56 postgraduate speciality training programs offers AI-related content [64]. To effectively integrate AI in medical education, it is essential to understand and address students' knowledge, skills and attitudes, as surveyed here, and actively involve students in curricular development. Educators should engage students at their individual learning levels while scientifically monitoring the integration process of AI in medical education to shape future guidelines [51,52,57]. A valid and very useful instrument to track the future effects of incorporating AI tools in medical curricula will be the ChatGPT literacy scale by Lee and Park (2024), that was unfortunately published only after this study was realised [65].

When developing AI teaching programs, it is crucial to enhance not only students' AI literacy but also their understanding of AI's future relevance in clinical practice. Such an approach, along with fostering a strong belief in AI's role in medicine, is key in motivating students' engagement and learning [66,67]. Clinically focused, small-group session can be an effective way to build interest in medical AI application [53]. Promoting "behavioral intention" in learning AI is especially valuable when tailored to specific students' subgroups, as analysed in this study. For example, medical curricula could target females who tend to feel more apprehensive towards AI, lower-performing students with less AI exposure, and those with limited prior knowledge who may undervalue AI's relevance. Furthermore, a flexible, voluntary online teaching format could support students' behavioural intentions by offering accessible, self-paced instructions – an ideal format to teach foundational topics in AI technology, e.g. machine learning. Given students' interest in applied AI skills, Lee et al., 2021, recommended a practical "hands-on" format to deepen engagement with AI tools.

Outlook

Before the release of ChatGPT in 2022, Masters highlighted AI's potential to reshape the role of future physicians and transform medical education [68]. Now, this potential is widely recognised by the public and students alike. While students are generally open to AI, they rate their own knowledge and skills as low, often using pragmatic, self-taught strategies to fill potential gaps. The fact that ChatGPT can pass medical exams has sparked a debate on current assessment methods [69]. However, it also shows that ChatGPT is currently still very much focused on basic knowledge reproduction, even though medical practice requires not only knowledge but also human interaction. As AI rapidly evolves, it is essential to rethink medical education to leverage the synergy between AI-supported learning and human-centred care [69].

Currently, universities have the chance to embrace AI as something more than only a novel teaching tool; it has the potential to revolutionise medical training. ChatGPT and similar AI applications could serve as "personalized digital learning companion", as

Balasooriya decribed it in 2024 [24,70]. To prepare students effectively, universities should define clear AI learning objectives and establish transparent guidelines for AI use in medical education, guided by facts on students' current skills and attitudes.

Acknowledgments

The authors thank the Faculty of Medicine's Teaching Research Group for their valuable input in designing this study and extend their gratitude to all survey participants.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

The author(s) reported there is no funding associated with the work featured in this article.

Notes on contributors

Lea Maaß is a dentist and third-year medical student at the Faculty of Medicine at Ulm University, Ulm, Germany. She is a doctoral student at the Institute of Anatomy and Cell Biology working on the topic presented in the present study for her thesis.

Claudia Grab-Kroll is an educational scientist and the head of the department of Studies and Teaching at the Faculty of Medicine at Ulm University, Ulm, Germany. She is responsible for coordinating and further developing the degree programs offered by the Faculty of Medicine. https://orcid.org/0000-0002-6688-2144

Jasmin Körner is a master of psychology and paedagogy and works as an instructional designer at the Competence Center Education in Medicine, Faculty of Medicine, Ulm, Germany. She specialises in the development of innovative digital teaching and learning concepts. Her research interests lie in particular in the teaching of digital literacies for students and teachers in the medical.field. https://orcid.org/0009-0004-1281-3356

Wolfgang Öchsner, MD, is senior anaesthetist at the Department of Anaesthesiology and Intensive Care Medicine at the University Hospital Ulm, Germany, with a specialisation in cardiac anaesthesiology. At the Faculty of Medicine of the Ulm University, he is head of the executive department for curricular development. https://orcid.org/0000-0001-7545-4777

Michael Schön, MD, is a physician scientist at the Institute of Anatomy and Cell Biology, Faculty of Medicine at Ulm University, Ulm, Germany. As a lecturer, he teaches medical and dental students in the gross anatomy course. His research focuses on changes at synapses related to autism spectrum disorders and neurodegenerative diseases. https://orcid.org/0000-0002-8905-1810

David Messerer, MD and associate professor at the Institute of Clinical Transfusion Medicine and Immunogenetics, German Red Cross Blood Transfusion Service, University Hospital, Ulm, Germany. Additionally, he qualified as a master of medical education. His research interest revolves around innovative approaches to diagnose and modulate excessive inflammation during sepsis and in medical education. https://orcid.org/0000-0002-0834-038X

Tobias Böckers, MD, is a professor and director of the Institute of Anatomy and Cell Biology, Faculty of Medicine at Ulm University, Ulm, Germany, where he also serves as the dean of studies. His research group investigates proteins and interactions of the ProSAP/Shank family related to post-synaptic density of synapses. https://orcid.org/0000-0002-1486-8535

Anja Böckers MD, is an associate professor and master of medical education at the Institute of Anatomy and Cell Biology, Faculty of Medicine at Ulm University, Ulm, Germany. As a senior lecturer, she teaches medical and dental students in the gross anatomy course and her research focus is on medical education related to anatomy competence. https://orcid.org/:0000-0002-2579-8632

ORCID

C. Grab-Kroll http://orcid.org/0000-0002-6688-2144
J. Koerner http://orcid.org/0009-0004-1281-3356
W. Öchsner http://orcid.org/0000-0001-7545-4777
M. Schön http://orcid.org/0000-0002-8905-1810
TM Böckers http://orcid.org/0000-0002-1486-8535
Anja Böckers http://orcid.org/0000-0002-2579-8632

References

[1] Gilson A, Safranek CW, Huang T, et al. U. a. How does ChatGPT perform on the United States medical licensing examination? The implications of large language models for medical education and knowledge assessment. JMIR Med Educ. 2023;9:e45312. doi: 10.2196/45312

[2] Kung TH, Cheatham M, Medenilla A, et al. U. a. Performance of ChatGPT on USMLE: potential for ai-assisted medical education using large language models. PloS Digit Health. 2023;2(2):e0000198. doi: 10.1371/journal.pdig.0000198

[3] Temsah O, Khan SA, Chaiah Y, et al. U. a. Overview of early ChatGPT's presence in medical literature: insights from a hybrid literature review by ChatGPT and human experts. Cureus URL cited 2023 Nov 12th]. Available from: https://www.cureus.com/articles/147971-overview-of-early-chatgpts-presence-in-medical-literature-insights-from-a-hybrid-literature-review-by-chatgpt-and-human-experts

[4] Neumann M, Rauschenberger M, Schön EM. "We need to talk about ChatGPT": The future of AI and Higher

education. In: 2023 IEEE/ACM 5th International Workshop on Software Engineering Education for the Next Generation (SEENG); Melbourne, Australia. IEEE; 2023 S. 29–32. URL cited 2023 Nov 12th]. Available from: https://ieeexplore.ieee.org/document/10190438/

[5] Tajik E, Tajik F. A comprehensive examination of the potential application of chat GPT in higher education institutions. URLcited 2023 Nov 12th]. Available from: https://www.techrxiv.org/articles/preprint/A_comprehensive_Examination_of_the_potential_application_of_Chat_GPT_in_Higher_Education_Institutions/22589497/1

[6] Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare. 2023;11(6):887. doi: 10.3390/healthcare11060887

[7] Sallam M, Salim N, Barakat M, et al. ChatGPT applications in medical, dental, pharmacy, and public health education: a descriptive study highlighting the advantages and limitations. Narra J. 2023;3(1):e103. doi: 10.52225/narra.v3i1.103

[8] Rasul T, Nair S, Kalendra D et al. u. a. The role of ChatGPT in higher education: benefits, challenges, and future research directions. J Appl Learn Teach. 2023;6(1). doi: 10.37074/jalt.2023.6.1.29

[9] Sok S, Heng K. ChatGPT for education and research: a review of benefits and risks. SSRN Electron J URL: https://www.ssrn.com/abstract=4378735 [cited 2023 Nov 12th]. doi: 10.2139/ssrn.4378735

[10] Fradkov AL. Early history of machine learning. IFAC-Pap. 2020;53(2):1385–1390. doi: 10.1016/j.ifacol.2020.12.1888

[11] Kaul V, Enslin S, Gross SA. History of artificial intelligence in medicine. Gastrointest Endosc. 2020;92(4):807–812. doi: 10.1016/j.gie.2020.06.040

[12] Weizenbaum J. Eliza—a computer program for the study of natural language communication between man and machine. Commun ACM. 1966;9(1):36–45. doi: 10.1145/365153.365168

[13] Thirunavukarasu AJ, Ting DSJ, Elangovan K, et al. Large language models in medicine. Nat Med. 2023;29(8):1930–1940. doi: 10.1038/s41591-023-02448-8

[14] Bassel A, Teixeira PEP, Pacheco-Barrios K, et al. Editorial - the use of large language models in science: opportunities and challenges. Princ Pract Clin Res J. 2023;9(1):1–4. doi: 10.21801/ppcrj.2023.91.1

[15] Hu K. Thomas Reuters. 2023 [cited 2023 Nov 20th]. Available from: https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/

[16] Dergaa I, Chamari K, Zmijewski P, et al. From human writing to artificial intelligence generated text: examining the prospects and potential threats of ChatGPT in academic writing. Biol Sport. 2023;40(2):615–622. doi: 10.5114/biolsport.2023.125623

[17] Sedaghat S. Early applications of ChatGPT in medical practice, education and research. Clin Med Mai. 2023;23(3):278–279. doi: 10.7861/clinmed.2023-0078

[18] Waisberg E, Ong J, Masalkhi M, et al GPT-4: a new era of artificial intelligence in medicine. Ir J Med Sci. 1971. doi: 10.1007/s11845-023-03377-8

[19] McKinney SM, Sieniek M, Godbole V, et al. u. a. International evaluation of an AI system for breast cancer screening. Nature. 2020;577(7788):89–94. doi: 10.1038/s41586-019-1799-6

[20] Rajpurkar P, Chen E, Banerjee O, et al. AI in health and medicine. Nat Med. 2022;28(1):31–38. doi: 10.1038/s41591-021-01614-0

[21] Artificial intelligence and machine learning (ai/ml)-enabled medical devices, US food & drug administration URL. [cited 2024 May 15th]. Available from: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices

[22] Bajaj T, Koyner JL. Artificial intelligence in acute kidney injury prediction. Adv Chronic Kidney Dis. 2022;29(5):450–460. doi: 10.1053/j.ackd.2022.07.009

[23] Lee S, Chu Y, Ryu J, et al. Artificial intelligence for detection of cardiovascular-related diseases from wearable devices: a systematic review and meta-analysis. Yonsei Med J. 2022;63(Suppl):93–107. doi: 10.3349/ymj.2022.63.s93

[24] Yu P, Xu H, Hu X, et al. Leveraging generative AI and Large language models: a comprehensive roadmap for healthcare integration. Healthcare Basel. 2023;11(20):2776. doi: 10.3390/healthcare11202776

[25] Wehkamp K, Krawczak M, Schreiber S. The quality and utility of artificial intelligence in patient care. Deutsches Ärzteblatt Int. 2023;120(27–28):463–469. Available from: https ://www .aerzteblatt.de/10 .3238/arztebl.m2023.0124

[26] Li J, Dada A, Kleesiek J, et al. ChatGPT in healthcare: a taxonomy and systematic review. Comput Methods Programs Biomed. 2024;245:108013. doi: 10.1016/j.cmpb.2024.108013

[27] Wang LKP, Paidisetty PS, Cano AM. The next paradigm shift? ChatGPT, artificial intelligence, and medical education. Med Teach. 2023;45(8):925. doi: 10.1080/0142159x.2023.2198663

[28] Lee H. The rise of ChatGPT: exploring its potential in medical education. Anat Sci Educ. 2023:ase.2270. doi: 10.1002/ase.2270

[29] Safranek CW, Sidamon-Eristoff AE, Gilson A, et al. The role of large language models in medical education: applications and implications. JMIR Med Educ. 2023;9:e50945. doi: 10.2196/50945

[30] Eysenbach G. The role of ChatGPT, generative language models, and artificial intelligence in medical education: a conversation with ChatGPT and a call for papers. JMIR Med Educ. 2023;9:e46885. doi: 10.2196/46885

[31] Lee J, Wu AS, Li D, et al. Artificial intelligence in undergraduate medical education: a scoping review. Acad Med. 2021;96(11S):S62–70. doi: 10.1097/acm.0000000000004291

[32] Hirosawa T, Shimizu T. Enhancing clinical reasoning with chat generative pre-trained transformer: a practical guide. Diagnosis. 2023 [cited 2024 May 15th]. doi: 10.1515/dx-2023-0116/html

[33] Zuckerman M, Flood R, Tan RJB, et al. u. a. ChatGPT for assessment writing. Med Teach 2. 2023 Nov;45(11):1224–1227. doi: 10.1080/0142159X.2023.2249239

[34] Ayers JW, Poliak A, Dredze M, et al. u. a. Comparing Physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med. 1. 2023 Juni 183(6):589. doi: 10.1001/jamainternmed.2023.1838

[35] Han Z, Battaglia F, Udaiyar A, et al. An explorative assessment of ChatGPT as an aid in medical education: use it with caution. Med Teach. 2023:1–8. doi: 10.1080/0142159x.2023.2271159

[36] Masters K. Ethical use of artificial intelligence in health professions education: AMEE guide No. 158. Med Teach. 2023;45(6):574–584. doi: 10.1080/0142159x.2023.2186203

[37] Tan LF, Heng JJY, Teo DB. Response to: „The next paradigm shift? ChatGPT, artificial intelligence, and medical education“. Med Teach. 2024;46(1):151–152. doi: 10.1080/0142159X.2023.2256961

[38] Paranjape K, Schinkel M, Nannan Panday R, et al. Introducing artificial intelligence training in medical education. JMIR Med Educ. 2019;5(2):e16048. doi: 10.2196/16048

[39] Redecker C, Punie Y. European framework for the digital competence of educators. DigCompedu. 2017. doi: 10.2760/178382 Joint Research Center Science for policy report of the European Commission.

[40] Schönbächler E, Himpsl-Gutermann K, Strasser T. Vom Chat zum Check. Informationskompetenz mit ChatGPT steigern. Medienimpulse. 2023. Bd. 61(1):51. doi: 10.21243/mi-01-23-18

[41] Laupichler MC, Aster A, Schirch J, et al. Artificial intelligence literacy in higher and adult education: a scoping literature review. Comput Educ Artif Intell. 2022;3:100101. doi: 10.1016/j.caeai.2022.100101

[42] Karaca O, Sa Ç, Demir K. Medical artificial intelligence readiness scale for medical students (MAIRS-MS) – development, validity and reliability study. BMC Med Educ. 2021;21(1):112. doi: 10.1186/s12909-021-02546-6

[43] Laupichler MC, Hadizadeh DR, Wintergerst MWM, et al. Adaption eines Instruments zur Erfassung der „AI Readiness“bei Medizinstudierenden. 2022. In: Annual Meeting of the Gesellschaft für medizinische Ausbildung (GMA); Halle. URL cited 2024 May 15th]. Available from: https://www.egms.de/en/meetings/gma2022/22gma187.shtml

[44] Boillat T, Nawaz FA, Rivas H. Readiness to embrace artificial intelligence among medical doctors and students: questionnaire-based study. JMIR Med Educ. 2022;8(2):e34973. doi: 10.2196/34973

[45] Civaner MM, Uncu Y, Bulut F, et al. Artificial intelligence in medical education: a cross-sectional needs assessment. BMC Med Educ. 2022;22(1):772. doi: 10.1186/s12909-022-03852-3

[46] Speidel R, Schneider A, Körner J, et al. Did video kill the XR star? Digital trends in medical education before and after the COVID-19 outbreak from the perspective of students and lecturers from the faculty of medicine at the University of Ulm. 2021;38(6): Doc 101 doi:10.3205/zma001497

[47] Körner J. Erfassung und Förderung digitaler Kompetenzen von Hochschullehrenden der Humanmedizin. Masterthesis, instructional design (M. Sc.), faculty of engineering, informatics and psychology, ulm university, Germany. 2024 [cited]. doi: 10.18725/OPARU-53128 [accessed June 27th.

[48] Kopfstatistik Statistik Sommersemester 2023. Universität Ulm; [cited 2024 Jun 27th]. https://www.uni-ulm.de/studium/studienorganisation/mehr/studierendenstatistik/statistik-sommersemester-2023

[49] Helfferich C. Leitfaden - und Experteninterviews. In: Baur N, Blasius J, editors. Handbuch Methoden der empirischen Sozialforschung. Springer Fachmedien Wiesbaden; 2014. doi: 10.1007/978-3-531-18939-0_39

[50] Wu M-J, Zhao K, Fils-Aime F. Response rates of online surveys in published research: a meta-analysis. Comput Hum Behav Rep. 2022;7:100206. doi: 10.1016/j.chbr.2022.100206

[51] Stewart J, Lu J, Gahungu N, et al. Western Australian medical students' attitudes towards artificial intelligence in healthcare. PLOS ONE. 2023 31;18(8):e0290642. doi: 10.1371/journal.pone.0290642

[52] Barreiro-Ares A, Morales-Santiago A, Sendra-Portero F, et al. Impact of the rise of artificial intelligence in radiology: what do students think? Int J Environ Res Public Health. 2023;20(2):1589. doi: 10.3390/ijerph20021589

[53] Blacketer C, Parnis R, Franke KB, et al. Medical student knowledge and critical appraisal of machine learning: a multicentre international cross-sectional study. Intern Med J. 2021;51(9):1539–1542. Available from: https://pubmed.ncbi.nlm.nih.gov/34541769/

[54] Pintos Dos Santos D, Giese D, Brodehl S, et al. Medical students' attitude towards artificial intelligence: a multicentre survey. Eur Radiol. 2019;29(4):1640–1646. doi: 10.1007/s00330-018-5601-1

[55] von Garrel J, Mayer J, Mühlfeld M. Künstliche Intelligenz im Studium. Eine quantitative Befragung von Studierenden zur Mutzung von ChatGPT & Co. Verlag Hochschule Darmstadt 2023. URL cited 2024 Jun 7th]. Available from: https://opus4.kobv.de/opus4-h-da/frontdoor/index/index/docId/395

[56] Fiebig R, Jeske L. Oral presentation. Tag der Lehre 19.10.2023, Stabsstelle Qualitätsentwicklung. Berichtswesen und Revision der Universität Ulm. [cited 2024 Jun 7th]. https://www.uni-ulm.de/einrichtungen/zentrum-fuer-lehrentwicklung/aktivitaeten/tag-der-lehre/tag-der-lehre-2023/#c970590

[57] Wong RS, Ming LC, Raja Ali RA. The intersection of ChatGPT, clinical medicine, and medical education. JMIR Med Educ. 2023;9:e47274. doi: 10.2196/47274

[58] Li YS, Lam CSN, See C. Using a machine learning architecture to create an AI-Powered chatbot for anatomy education. Med Sci Educ. 2021;31(6):1729–1730. doi: 10.1007/s40670-021-01405-9

[59] Ayre J, Mac O, McCaffery K, et al. New frontiers in health literacy: using ChatGPT to simplify health information for people in the community. J Gen Intern Med. 2024 Mar;39(4):573–577. doi: 10.1007/s11606-023-08469-w

[60] Cox A, Seth I, Xie Y, et al. Utilizing ChatGPT-4 for providing medical information on blepharoplasties to patients. Aesthet Surg J. 2023 Jul 15;43(8):NP658–NP662. doi: 10.1093/asj/sjad096

[61] 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(4):446–470. doi: 10.1080/0142159x.2024.2314198

[62] Nationaler Kompetenzbasierter Lernzielkatalog Medizin (NKLM) Version 2.0. des Medizinischen Fakultäten Tages. Berlin; 2021.URL cited 2024 Jun 20th]. Available from: https://nklm.de/zend/menu

[63] Mosch L, Agha-Mir-Salim L, Sarica MM, et al. Artificial intelligence in undergraduate medical education. Stud Health Technol Inform. 2022;294:821–822. doi: 10.3233/shti220597
[64] Agha-Mir-Salim L, Mosch L, Klopfenstein SAI, et al. Artificial intelligence competencies in postgraduate medical training in Germany. Stud Health Technol Inform. 2022;294:805–806. doi: 10.3233/shti220589
[65] Lee S, Park G. Development and validation of ChatGPT literacy scale. Curr Psych. 2024;43(21):18992–19004. doi: 10.1007/s12144-024-05723-0
[66] Li X, Jiang MY, Jong MS, et al. Understanding medical students' perceptions of and behavioral intentions toward learning artificial intelligence: a survey study. Int J Environ Res Public Health. 2022;19(14):8733. doi: 10.3390/ijerph19148733
[67] Wagner G, Raymond L, Paré G. Understanding prospective physicians' intention to use artificial intelligence in their future medical practice: configurational analysis. JMIR Med Educ. 2023;9:e45631. doi: 10.2196/45631
[68] Masters K. Artificial intelligence in medical education. Med Teach. 2019;41(9):976–980. doi: 10.1080/0142159x.2019.1595557
[69] Mbakwe AB, Lourentzou I, Celi LA, et al. ChatGPT passing USMLE shines a spotlight on the flaws of medical education. PloS Digit Health. 2023;2(2):e0000205. doi: 10.1371/journal.pdig.0000205
[70] Balasooriya C. AI in education: a futuristic vision. Med Teach. 2024;9:1–3. doi: 10.1080/0142159x.2024.2352160