Dermatology Reports
https://journals.pagepress.net/dr

eISSN 2036-7406

Please cite this article as:

Naldi L, Valetto MR, Cazzaniga S, et al. "Acne in the Age of ChatGPT": an artificial intelligence-generated and evidence-based continuing medical education distance course for Italian physicians. *Dermatol Rep 2026 [Epub Ahead of Print] doi: 10.4081/dr.2026.10935*

Received: 11 May 2026; Accepted: 14 August 2026.

"Acne in the Age of ChatGPT": an artificial intelligence-generated and evidence-based continuing medical education distance course for Italian physicians

Luigi Naldi,[1,2] Maria Rosa Valetto,[3] Simone Cazzaniga,[1,4] Sergio Cima,[3] Silvia Emendi,[3] Nicoletta Scarpa,[3] Eugenio Santoro,[5] Pietro Dri,[3] Vincenzo Bettoli[1,6]

[1]Center of the Italian Group for Epidemiological Research in Dermatology, Bergamo, Italy; [2]Dermatology Unit, Ospedale San Bortolo, Vicenza, Italy; [3]Zadig Ltd Benefit Company, CME National Provider, Milan, Italy; [4]Department of Dermatology, Inselspital University Hospital of Bern, Switzerland; [5]Unit of Research in Digital Health and Digital Therapeutics, Department of Clinical Oncology, Mario Negri Institute for Pharmacological Research, Milan, Italy; [6]Section of Dermatology and Infectious Diseases, Department of Medical Sciences, University of Ferrara, Italy

Correspondence: Maria Rosa Valetto, Zadig Ltd Benefit Company, via Ampère, 59, 20131, Milan, Italy. E-mail valetto@zadig.it

Key words: acne; artificial intelligence; large language models; continuing medical education; evidence-based medicine.

Contributions: Vincenzo Bettoli, Pietro Dri, Luigi Naldi, Eugenio Santoro, Maria Rosa Valetto: conceptualization and methodology development; Sergio Cima, Silvia Emendi, Nicoletta Scarpa, Maria Rosa Valetto: data curation; Simone Cazzaniga, Sergio Cima, Maria Rosa Valetto: formal analysis, visualization; Pietro Dri, Silvia Emendi, Luigi Naldi, Nicoletta Scarpa, Maria Rosa Valetto, Vincenzo Bettoli, Eugenio Santoro, validation; Maria Rosa Valetto: writing – original draft; Vincenzo Bettoli, Simone Cazzaniga, Luigi Naldi, Luigi Naldi: writing – review & editing; Vincenzo Bettoli, Pietro Dri: supervision. All authors read and approved the final version to be published.

Conflict of interest: MRV, SCi, SE, NS, and PD are employees of Zadig Ltd Benefit Company, the CME National Provider responsible for developing and hosting the e-learning course on the SAEPE platform. The authors declare that the study was conducted in accordance with the principles of transparency and scientific integrity. The other author have no conflict of interest to declare.

Ethics approval and consent to participate: all procedures contributing to this study comply with the ethical standards of the relevant national and institutional committees on medical research and

with the Helsinki Declaration of 1975, as revised in 2024. Ethics approval is not required by regulatory authorities due to: i) the nature of the study (analysis of the participation and satisfaction data about an e-learning course); ii) the full anonymization of the study participants who agreed, by signing an online authorization, that anonymized data on their learning activity would be used for statistical purposes; and iii) the voluntary enrollment of health professionals and not of patients, without any collection of health-related data.

Availability of data and materials: the data supporting the findings of this study are available in the supplementary files of this article or from the corresponding author upon reasonable request.

Funding: this research received no grant from any funding agency in the public, commercial, or not-for-profit sectors.

Declaration of generative artificial intelligence and artificial intelligence-assisted technologies in the writing process: the authors declare that no generative AI tools were used to assist in drafting, editing, or translating the text of this manuscript.

Abstract

The AI-CHECK (Artificial Intelligence for CME Health E-learning Contents and Knowledge) study aimed to explore ChatGPT's potential in continuing medical education (CME). In particular, its third phase investigated artificial intelligence (AI)'s ability to develop a CME course on acne in Italian and assessed participants' satisfaction and attitudes toward AI-generated content. The e-learning course "Acne in the Age of ChatGPT" was hosted on the SAEPE platform and offered free to Italian physicians from May to September 2025. It included a handbook, two clinical vignettes, a knowledge questionnaire, and pre- and post-course surveys. ChatGPT-4 generated all educational materials following structured prompts. Content was reviewed by an expert dermatologist. A total of 556 physicians enrolled, with 95% completing the course. Most participants (74.4%) expressed support for AI use in medicine, while 79.8% recognized its value in developing CME materials. Overall satisfaction was excellent: over 95% rated the course as relevant, high-quality, and effective. This study demonstrates that ChatGPT can generate high-quality CME content with human supervision, ensuring accuracy and reliability. AI integration in CME shows promise for scalable, personalized, and innovative medical education.

Introduction

Among the wide range of applications of artificial intelligence (AI) in medicine,[1] large language models (LLMs) hold a significant role in the training of health-care professionals, with potential impact in several important areas. They can offer support for educators in developing instructional plans and both academic and postgraduate training content, generating assessment questions to evaluate knowledge and skills, creating interactive simulated clinical scenarios to strengthen decision-making abilities, supporting medical-patient communication through realistic dialogue generation, and contributing to the creation of interactive educational resources.[2-7]

However, challenges persist, including risks of academic misconduct, over-reliance on AI, potential dilution of critical thinking skills, and concerns about content accuracy and reliability.[1,8-10]

The AI-CHECK (Artificial Intelligence for CME Health E-learning Contents and Knowledge) study, focusing on acne, was designed in three steps to explore the potential of ChatGPT in continuous medical education (CME) in dermatology. In the first step,[11] we analyzed the strengths and limitations of ChatGPT in providing information on acne to the general population. In the second step,[12] we evaluated the materials produced by ChatGPT for a CME course targeting general practitioners and compared them with the recommendations of the recent National Institute for Health and Care Excellence (NICE) guidelines.[13]

Here, we present the results of the third and final step of the AI-CHECK study, which explored the capabilities and performance of ChatGPT in producing a comprehensive and evidence-based distance learning CME course on acne, as well as the level of participants' satisfaction. Participants were also asked to complete a structured questionnaire designed to assess their attitudes toward AI and the perceived quality of the ChatGPT-generated educational content.

Materials and Methods

Structure and contents of the e-learning course

The e-learning course titled "Acne in the Age of ChatGPT" was offered free of charge for 4 months (May 7-September 7, 2025) to the Italian physicians registered to the e-platform SAEPE (more than 43,000 as of November 2025; www.saepe.it), which is based on the open-source Moodle Learning Management System and since 2009 has hosted more than 730 CME distance courses targeted to healthcare professionals. To earn the 5 CME credits, the participants had to answer at least 75% of questions correctly.

It was structured in compliance with the regulations of the Italian Continuing Medical Education system and included a handbook (see an extract in the *Supplementary Table 1*), two clinical vignettes (see an example in the *Supplementary Table 2*), and a CME knowledge questionnaire.

The handbook was created by assembling the series of questions and answers produced during the second phase of the AI-Check study[12] when ChatGPT version 4 (released on March 14, 2023) was exposed to a 23-item questionnaire addressing the main issues in managing acne (*Supplementary Table 3*). Clinical vignettes, questionnaire, and corresponding answers were generated by ChatGPT following detailed instructions provided by prompts (*Supplementary Table 4*). For the purposes of this third phase of the AI-CHECK study, ChatGPT was synchronously exposed to the 23-item questionnaire in both English and Italian, and a comparison between the two sets of answers was performed. This comparison was necessary to confirm that the conclusions from phases 1 and 2[11,12] of the AI-CHECK study on the validity of ChatGPT-generated content in English could be extended to content generated in Italian, the language used in the CME course.

Clinical vignettes were clinical scenarios portrayed by imaginary actors in five narrative steps, intercalated with five multiple-choice questions focused on decision-making and problem-solving, each with four answers, of which only one was correct.

The CME knowledge questionnaire consisted of 30 questions, each with four answers, of which only one was correct.

Revision of the course contents

Before publication, the distance learning course was reviewed by an international expert on acne (VB) who, as scientific lead of the course, amended and supplemented the information provided by ChatGPT, while clearly indicating any changes so that participants could understand what had been done by AI and what had been done by the reviewer. Considering the handbook and the vignettes, the reviewer's changes and integrations were placed in square brackets within the text, while the CME questions were accepted or rejected without any tracking. Changes were aimed at improving clinical consistency and accuracy of the scientific information and at ensuring clarity of questions and answers.

Assessment of participants' satisfaction about the course and attitude towards AI

Participants' satisfaction was assessed by means of the standardized, anonymous CME system satisfaction questionnaire, mandatory under Italian CME regulations. Moreover, a survey on participants' attitudes towards AI and the perceived quality of ChatGPT-generated content was developed. It consisted of two parts, one administered before and one after the course (*Supplementary Table 5*). The pre-course section investigated participants' knowledge, experiences, and opinions on generative AI in medicine, including perceived benefits and risks. The post-course section assessed satisfaction with the course content and materials, evaluating clarity, relevance, quality, and the perceived contribution of AI to educational quality, with optional open-ended feedback (*Supplementary Table 6*).

Text comparison

Text comparisons were conducted on two separate occasions: the first aimed to compare the English and Italian responses generated by ChatGPT, and the second to evaluate the modifications introduced by the reviewer in the handbook and vignettes with respect to the original version derived from ChatGPT's outputs.

Specifically, a pipeline designed for domain-specific, cross-lingual analysis was used:[14]

i) Specialized embedding generation: The Clinical BERT model (emilyalsentzer/Bio_ClinicalBERT) was employed to convert each text into a high-dimensional vector representation (embedding). This model is specifically pre-trained on biomedical literature, enhancing sensitivity to clinical terminology.

ii) Similarity measurement: the cosine similarity was computed between all resulting vectors. Cosine similarity quantifies the angular difference between two vectors; a value of 1 indicates identical meaning, while 0 indicates no shared meaning. Negative values may occur, representing opposite or contrasting meanings.

In the first comparison, which involved two different languages, this pipeline was preceded by a cross-lingual normalization step to ensure language uniformity across the dataset: the Italian texts were automatically translated into English using the googletrans library.

Statistical analysis

Descriptive statistics were computed via spreadsheet application (Excel, Microsoft Windows®), and analyses were performed with R software version 4.3.3 (R Foundation for Statistical Computing, Vienna, Austria).

The non-parametric Mann-Whitney U test for independent samples was used to compare participants' characteristics with their evaluation of the AI-developed educational materials, specifically those related to this course. Statistical significance was set at $p<0.05$.

Results

Comparison of AI-generated content in English vs. *Italian languages*

An average pairwise similarity of approximately 98% (mean cosine similarity score 0.99, 95% confidence interval [CI]: 0.98-0.99) across the two sets of answers was estimated, indicating a state of near-semantic equivalence. This result implies that the core clinical information and specialized terminology – as encoded by ClinicalBERT – is highly consistent. Thus, Italian content could be considered equivalent to the English content in terms of the characteristics explored in phases 1 and 2 of the AI-CHECK study.[11,12]

Nature and extent of changes in the text of the course materials

The comparison between the original and final Italian text of the handbook shows that the revised version underwent moderate structural changes (*Supplementary Table 7*) and maintained a high similarity (cosine similarity score 0.99, 98.7%).

Similarly, after the reviewer intervention, the vignettes underwent moderate structural changes while similarity remained extremely high, with a cosine similarity score of 0.97 (97.00%) for vignette 1 and of 0.97 (97.44%) for vignette 2.

Of the 30 CME questions generated by ChatGPT, 6 (20%) were deleted because they did not comply with the prompted instructions or required substantial modifications, while the remaining 24 were used in the original version proposed by ChatGPT.

E-learning course participation and achievement

Over 4 months, 556 physicians (Table 1) enrolled for the course in all Italian regions.

Overall, 528 (95%) of the health professionals earned CME credits, while 28 (5%) did not complete the course within the established time frame.

Satisfaction questionnaire and survey

The anonymous customer questionnaire at the end of the course provided highly positive judgments: 95.4% rated the course as relevant, 96% rated it as of quality, and 96% rated it as effective (Figure 1). Also, 69 of those who completed the course (13%) wrote open-ended feedback in the evaluation form. The message was positive in 85.5% of cases; "interesting" (n=14, 20%) was the most frequently extrapolated word from the text, followed by "useful" (n=13, 18.8%) and "good" (n=10, 14.4%) (*Supplementary Table 6*).

A higher number of participants (n=632) completed the pre-course section of the survey compared to those who engaged in the course up to the CME questionnaire. The results of the survey are presented in detail in *Supplementary Table 5*. In the pre-course section, the majority of participants (55.9%) indicated awareness of generative AI, but 31% reported that they had never used it.

The most common declared occasions for using AI in the medical field were to obtain information about diseases (41.7%) and medications (34.7%). Additionally, AI was used for bibliographic searches (19.7%), text translation (19.5%), analysis and synthesis of documents (18.7%), diagnostic support (18.4%), and preparation of presentations (12.4%).

Most participants (74.3%) supported using AI in medicine, while 23.6% were doubtful. On the other hand, 56% considered its use risky or problematic, while 30.7% had no clear position. The greatest benefit identified was access to scientific information (60%), while the greatest risk was the replacement of the physician's role (39.2%).

In the pre-course questionnaire, 54.3% of respondents stated that AI can be used to generate high-quality CME courses, while 51.9% indicated that it can be risky to delegate the autonomous production of CME courses to generative AI.

After completing the course, in the post-course questionnaire, a high proportion of participants (79.8%) expressed the view that generative AI can play a valuable role in the development of high-quality training materials for CME. However, it is notable that 74.8% of respondents also emphasized the importance of transparent use of AI by the provider.

Most participants rated the quality of the handbook, vignettes, and questionnaire positively.

Regarding the perceived contribution of AI to the development of medical education courses (Question 1.1 of the post-course section of the survey), participants with previous experience in generative AI expressed greater trust compared with those without such experience (experienced *vs*. naïve, grouped according to Question 1.2 of the pre-course section of the survey, $p<0.001$). Similarly, participants

with a favorable attitude towards AI showed higher trust than those who were skeptical or neutral (enthusiast *vs.* skeptic/neutral, grouped according to Question 2.1 of the pre-course section of the survey, $p<0.001$) (*Supplementary Table 8*).

Regarding the evaluation of course materials (*Supplementary Table 9* and *Supplementary Figure 1*) explored in the post-course sections 2, 3, and 4 of the survey, no significant differences were observed between experienced and naïve participants in their evaluation of the handbook in terms of clarity, completeness, organization, reliability, or up-to-dateness. By contrast, experienced participants expressed significantly more favorable opinions of the AI-generated vignettes in terms of clarity ($p<0.05$), narrative quality ($p<0.01$), plausibility ($p<0.05$), ability to explore clinical aspects ($p<0.05$), and wording ($p<0.01$). They also rated the CME knowledge questionnaire more positively in terms of clarity of the answers ($p<0.05$), ability to explore clinical aspects ($p<0.05$), and plausibility ($p<0.05$). On the other hand, when comparing participants with different attitudes towards AI, the enthusiastic participants consistently expressed higher satisfaction with the handbook, vignettes, and CME knowledge questionnaire, with all comparisons reaching statistical significance ($p<0.05$ to $p<0.001$).

Discussion

To the best of our knowledge, AI-CHECK is the first research project describing the development and delivery of a CME course, compliant with Italian regulatory requirements, entirely generated by AI. Our results show that this kind of course meets the interest and satisfaction of its participants and suggest that AI can assist in the creation of a CME course, but it cannot be relied upon to produce an entire course, as human supervision is still essential to avoid errors and inappropriate information.

We are confident that our study design reflects careful attention to the broader principles of AI governance in medical education – namely expert supervision, accountability, transparency, and quality assurance – and we regard this as one of the strengths of our work. In keeping with these principles, all AI-generated materials were reviewed by an internationally recognized expert in acne, who verified the clinical consistency, accuracy, and adherence to evidence-based medicine before dissemination to learners. Responsibility for the accuracy and reliability of the content rested with the human reviewer and the course provider, not the AI system, ensuring clear accountability for any errors or inaccuracies. In line with a transparent approach to the use of AI, learners were also explicitly informed, in the course introduction published on the SAEPE platform, that the educational content had been AI-generated, together with an explanation of the generation process itself, the review methodology, and the modality used to flag the edits. Finally, quality assurance was embedded in the structured content production process described in the Materials and Methods and further detailed in the *Supplementary Material*, with a clear separation of roles between automated content generation

(ChatGPT-4, *via* structured prompts) and independent scientific validation; it should also be noted that Zadig (www.zadig.it), the course provider, holds UNI EN ISO 9001:2015 quality certification for distance learning. The approach adopted in our study could and should be further refined and systematically adopted whenever AI is used to produce educational material.

The development of our course represents the third phase of a research project whose precedent phases of validation of ChatGPT-generated content on acne were conducted entirely in English. Given that the literature provides no conclusive evidence on ChatGPT's performance in languages other than English, including Italian,[15-18] it was important to verify and confirm that the conclusions drawn from the English texts were also applicable when querying ChatGPT in Italian.

It should be emphasized that we aimed to reproduce, with an LLM such as ChatGPT, a course format that has been extensively tested in our experience as a provider,[19] and for which evidence of efficacy is available.[20-22] Our experience, never tried before, contributes to the existing evidence on the applications of LLMs such as ChatGPT in postgraduate medical education, more specifically in CME programs.

One notable strength of LLMs is their capacity to generate realistic and diverse clinical scenarios, which are essential for preparing physicians. This is relevant as previous studies[20-22] have shown that internet-based CME centered on clinical vignettes can lead to measurable changes in physician behavior as well as sustained gains in knowledge.[23] Moreover, physicians who participate in this type of e-learning activity are more likely to make evidence-based clinical decisions compared with those who do not participate.[24] Accordingly, ChatGPT has demonstrated the capability to generate role-play scripts for plausible clinical situations.[25]

ChatGPT has been shown to be capable of rapidly generating realistic clinical scenarios, including in the field of dermatology; however, expert review remains crucial to ensure accuracy and alignment with clinical and educational standards.[26,27]

In the AI-CHECK study, ChatGPT performed adequately in implementing clinical vignettes with intercalated CME questions, which are the key feature of the educational model adopted.

Additionally, LLMs demonstrate proficiency in creating quiz questions and assessments that are comparable in quality and readability to those authored by human experts.[5,8,28] These findings are consistent with our experience when generating the CME questionnaire.

Nonetheless, some limitations and errors in ChatGPT-derived materials required correction by the reviewer, including aspects of acne management, *e.g.*, incorrect guidance on antibiotics, insufficient focus on treatment safety and side effects, and inaccuracies regarding long-term outcomes. In addition, certain statements exhibited issues with tone, such as employing terms like "might" when a more assertive approach was warranted, or, conversely, failing to adequately express necessary uncertainty.

Overall, these observations underscore the substantial risks associated with relying exclusively on artificial intelligence for medical education. They highlight the critical importance of thorough expert review to ensure that misinformation is not inadvertently circulated.

The participation in this educational offering can be considered more than satisfactory, exceeding that recorded among more than 1,800 physicians attending courses in 2025 on the SAEPE platform for 66 comparable courses with the same number of CME credits (mean participation: 29; range: 4-237). It can be hypothesized that the novelty of an AI-based course, together with the carefully explained methodology presented in the course introduction, sparked the interest or curiosity of participants. On the other hand, the fact that the educational program was offered free of charge may have encouraged participation.

A positive attitude among participants, reflecting curiosity and critical engagement, also emerges from the responses to questionnaires. Attitudes were cautiously positive: AI was perceived as useful and potentially innovative, yet concerns about accuracy, ethics, and impact on human roles were evident. The usage frequency and applications are in line with previous data.[29]

Since participation in the course was voluntary, a potential self-selection bias cannot be excluded, as physicians more interested in AI might have been more inclined to enroll and to provide favorable evaluations. However, the pre-course questionnaire (*Supplementary Table 5*) showed that just over half of participants reported knowing what generative artificial intelligence is, while only slightly more than one-fifth reported making intensive use of it (daily/weekly); this distribution does not seem to indicate a marked skew of participants toward heavy AI users, suggesting that, if present, self-selection bias was unlikely to have substantially influenced the results.

The level of satisfaction, assessed in terms of the relevance and quality of the content and the effectiveness of the training, was in line with that reported on the SAEPE platform for other courses (always >95% when grouping the top three levels of the Likert scale). However, a few concerns emerged about the ability of AI to fully replace traditional teaching. Overall, the feedback suggests that AI-produced courses are positively received, while leaving space for improvement and integration with human guidance.

We documented that prior exposure to generative AI and a favorable attitude towards its use contribute to greater confidence in its potential role in medical education.

Although assessing knowledge and/or competence acquisition was not an objective of the present study, it is worth noting that the course pass threshold, set at 75% and achieved by 95% of participants, indirectly suggests a satisfactory level of proficiency in the medical topic covered.

One limitation of our study is that we did not apply a trial design by subjecting two groups of participants to a course authored entirely by AI or by human experts.[4,30] If this model had been

followed, it would not have been possible to administer an accredited CME course because there would have been no prior control of the content and quality of the course written entirely by AI, exposing learners to misinformation.

Another possible limitation is that the course was provided free of charge, which may have encouraged positive comments rather than criticism.

Finally, this study reports experience from a single topic (acne), one national CME platform, and one LLM (ChatGPT-4); accordingly, the generalizability of the results remains to be confirmed by replicating the study across different educational settings or AI tools.

Conclusions

The AI-CHECK study shows that large language models like ChatGPT can effectively contribute to the creation of high-quality, engaging CME content, fostering innovation in medical education, and able to meet interest and satisfaction of participants. However, as several inaccuracies, even scientifically relevant ones, can occur, expert oversight is needed to ensure scientific validity and educational reliability.

References

1. European Commission: Directorate-General for Health and Food Safety, EEIG, Open Evidence and PwC, Study on the deployment of AI in healthcare – Final report, Publications Office of the European Union, 2025. Available from: https://data.europa.eu/doi/10.2875/2169577
2. Safranek CW, Sidamon-Eristoff AE, Gilson A, Chartash D. The role of large language models in medical education: applications and implications. JMIR Med Educ 2023;9:e50945.
3. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare (Basel) 2023;11:887.
4. Lavigne E, Lopez A, Frandon J, et al. AI-standardized clinical examination training on OSCE performance. NEJM AI 2025;2.
5. Li R, Wu T. Delving into the practical applications and pitfalls of Large Language Models in medical education: Narrative Review. Adv Med Educ Pract 2025;16:625-36.
6. Lucas HC, Upperman JS, Robinson JR. A systematic review of large language models and their implications in medical education. Med Educ 2024;58:1276-85.
7. Gracey LE, Nambudiri VE, Rotemberg V, Maderal AD. Challenges and opportunities for AI in dermatology residency. JAMA Dermatol 2025;161:1103-4.
8. Benítez TM, Xu Y, Boudreau JD, et al. Harnessing the potential of large language models in medical education: promise and pitfalls. J Am Med Inform Assoc 2024;31:776-83.

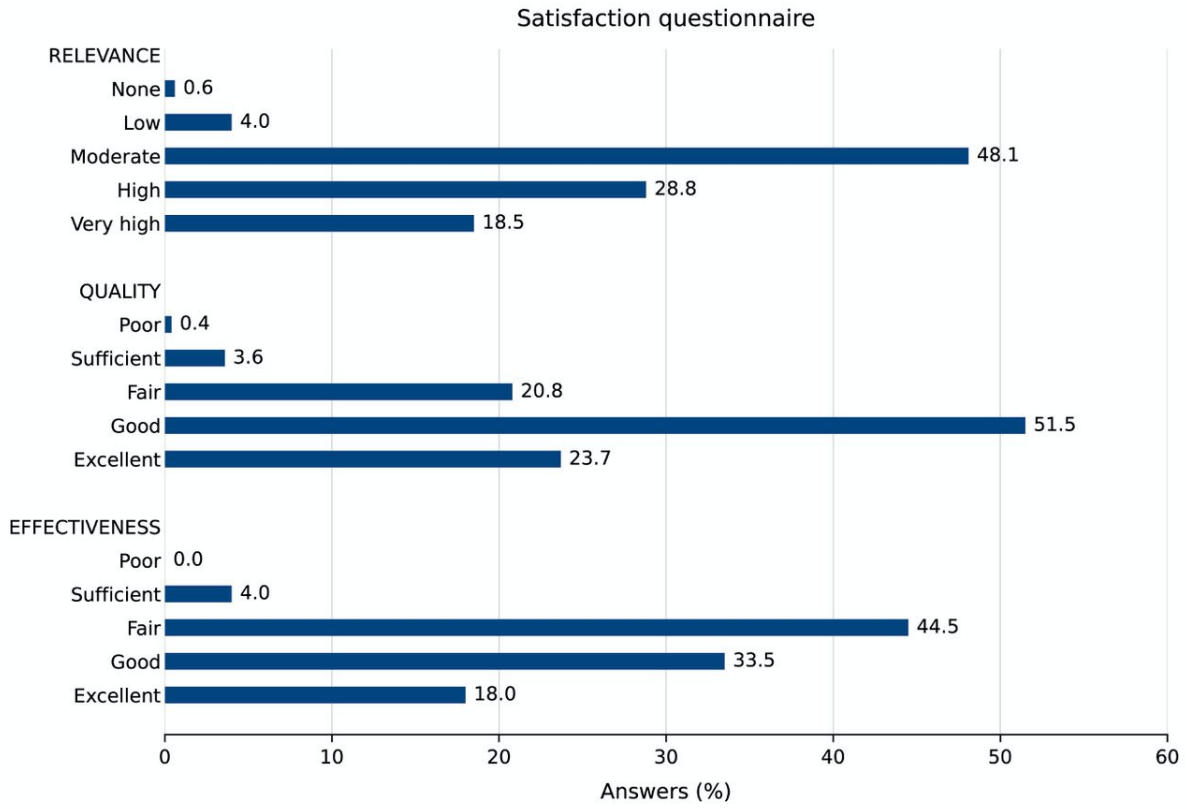
9. Aydin S, Karabacak M, Vlachos V, Margetis K. Large language models in patient education: a scoping review of applications in medicine. Front Med (Lausanne) 2024;11:1477898.
10. Meskó B, Topol EJ. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. NPJ Digit Med 2023;6:120.
11. Bettoli V, Naldi L, Santoro E, et al. ChatGPT and acne: Accuracy and reliability of the information provided - The AI-check study. J Eur Acad Dermatol Venereol 2025;39:e359-62.
12. Naldi L, Bettoli V, Santoro E, et al. Application of ChatGPT as a content generation tool in continuing medical education: acne as a test topic. Dermatol Reports 2025;17:10138.
13. National Institute for Health and Care Excellence (NICE). Acne vulgaris: management. NICE guideline 198; 2021. Available from: https://www.nice.org.uk/guidance/ng198
14. Kishore MV, Bodapati P. High-performance semantic similarity analysis for medical research documents using transformer models (BioBERT/ClinicalBERT) with WMD/WMS. J Theor Appl Inf Technol 2025;103:2842-56.
15. Fang C, Wu Y, Fu W, et al. How does ChatGPT-4 preform on non-English national medical licensing examination? An evaluation in Chinese language. PLOS Digit Health 2023;2:e0000397.
16. Zhang X, Li S, Hauer B, et al. Don't trust ChatGPT when your question is not in English: a study of multilingual abilities and types of LLMs. arXiv [Preprint]. 2023.
17. García-Ferrero I, Agerri R, Salazar Atutxa A, et al. MedMT5: an open-source multilingual text-to-text LLM for the medical domain. arXiv [Preprint]. 2024.
18. Ray M, Kats DJ, Moorkens J, et al. Evaluating a Large Language Model in translating patient instructions to Spanish using a standardized framework. JAMA Pediatr 2025;179:1026-33.
19. Moja L, Moschetti I, Liberati A, et al. Clinical Evidence and its use in a National Continuing Medical Education Program in Italy. Plos Medicine 2007;4:e113.
20. Cervero RM, Gaines JK. The impact of CME on physician performance and patient health outcomes: an updated synthesis of systematic reviews. J Contin Educ Health Prof 2015;35:131-8.
21. Noesgaard SS, Ørngreen R. The effectiveness of e-learning: an explorative and integrative review of the definitions, methodologies and factors that promote e-Learning effectiveness. Electron J e-Learning 2015;13:278-90.
22. Vaona A, Banzi R, Kwag KH, et al. E-learning for health professionals. Cochrane Database Syst Rev 2018;1:CD011736.
23. Fordis M, King JE, Ballantyne CM, et al. Comparison of the instructional efficacy of internet-based CME with live interactive CME workshops: A randomized controlled trial. JAMA 2005;294:1043-51.

24. Casebeer L, Engler S, Bennett N, et al. A controlled trial of the effectiveness of internet continuing medical education. BMC Med 2008;6:37.
25. Scherr R, Halaseh FF, Spina A, et al. ChatGPT interactive medical simulations for early clinical education: Case study. JMIR Med Educ 2023;9:e49877.
26. Ghaffari F, Langarizadeh M, Nabovati E, Sabery M. Effectiveness of ChatGPT for clinical scenario generation: A qualitative study. Arch Acad Emerg Med 2025;13:e49.
27. Dunn C, Hunter J, Steffes W, et al. Artificial intelligence-derived dermatology case reports are indistinguishable from those written by humans: A single-blinded observer study. J Am Acad Dermatol 2023;89:388-90.
28. Laupichler MC, Rother JF, Grunwald Kadow IC, et al. Large Language Models in medical education: Comparing ChatGPT- to human-generated exam questions. Acad Med 2024;99:508-12.
29. Blease CR, Locher C, Gaab J, et al. Generative artificial intelligence in primary care: an online survey of UK general practitioners. BMJ Health Care Inform 2024;31:e101102.
30. Rao A, Artino A. AI-Driven OSCE preparation in medical education: promise, pitfalls, and practical implications. NEJM AI 2025;2.

Table 1. Profile of the physicians enrolled in the CME course "Acne at the Age of ChatGPT".

Participant characteristics	**Values (N=556)**
Age	
Mean±SD	52.4±15.1 years
Median	54 years
Range	25-80 years
Sex	
Females	262 (47.1%)
Males	293 (52.7%)
Non-binary	1 (0.2%)
Residence	
Northern Italy	299 (53.7%)
Central Italy	144 (25.9%)
Southern Italy and Islands	113 (20.4%)
Most frequent professional profile	General practitioners: 131 (23.6%)
	Other specialists: 385 (69.2%)
	Non specialists: 40 (7.2%)

Figure 1. Results of the satisfaction questionnaire.



Online Supplementary Material:

Supplementary Table 1. *Extract from the handbook.*

Supplementary Table 2. *Example of the clinical vignette.*

Supplementary Table 3. *List of the 23 questions prompted to ChatGPT 4.0.*

Supplementary Table 4. *Prompts used to generate the contents (handbook, vignettes, and questionnaire) of the CME course "Acne at the age of ChatGPT".*

Supplementary Table 5. *Survey on participants' attitudes towards AI.*

Supplementary Table 6. *Open-ended comments provided by the course participants.*

Supplementary Table 7. *Analysis of structural changes in the handbook and the two clinical vignettes.*

Supplementary Table 8. *Questions post-course, in total and by comparison groups.*

Supplementary Table 9. *Questions about the course materials, in total and by group.*

Supplementary Figure 1. Questions about the course materials in total and by group. Statistically significant differences are shown as * for $p<0.05$, ** for $p<0.01$, and *** for $p<0.001$.