



## Is reflective writing a valid summative assessment in the era of GenAI? A scoping review

Emma Furzer & Philip M. Newton

To cite this article: Emma Furzer & Philip M. Newton (13 May 2026): Is reflective writing a valid summative assessment in the era of GenAI? A scoping review, *Medical Teacher*, DOI: [10.1080/0142159X.2026.2658172](https://doi.org/10.1080/0142159X.2026.2658172)

To link to this article: <https://doi.org/10.1080/0142159X.2026.2658172>



© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Published online: 13 May 2026.



[Submit your article to this journal](#)



Article views: 160



[View related articles](#)



[View Crossmark data](#)

# Is reflective writing a valid summative assessment in the era of GenAI? A scoping review

Emma Furzer  and Philip M. Newton 

Swansea University, Swansea, UK

## ABSTRACT

**Introduction:** Reflective practice is central to medical education at all levels and is commonly assessed using asynchronous written reflections. The emergence of new generative artificial intelligence tools such as ChatGPT has led to concern that they may be used to generate written reflections and so to bypass the reflective process, potentially undermining the value of written reflections as a summative assessment. This scoping review asked whether written reflections remain a valid summative assessment in the age of GenAI by identifying papers which examined the ability of GenAI to write reflectively, the detection of such writing, and the use of GenAI for reflection by health professionals.

**Methods:** A scoping review was conducted in accordance with PRISMA-ScR guidelines using Medline, CINAHL, Education Research Complete and Google Scholar. Data were charted and the descriptive content thematically analysed.

**Results:** Eighteen publications were included. We find that free GenAI chatbots such as ChatGPT can write reflections to a high academic standard, and these are difficult to detect either by humans or software. We also find that health professionals and students are using GenAI for this purpose, but also that GenAI might be used to support and enhance the reflective process through scaffolding, acting as a conversational partner and assisting in the redesigning of assessments.

**Discussion:** The findings suggest that asynchronous written reflection is vulnerable as a summative assessment in the presence of GenAI. Future research should seek to answer the question as to how the validity of reflective writing as a summative assessment can be maintained in a world where GenAI use continues to grow.

## ARTICLE HISTORY

Received 28 January 2026  
Accepted 7 April 2026

## KEYWORDS

Reflective practice; ChatGPT; assessment; GenAI; health professions education

## Introduction

Reflection can be defined as ‘the process whereby an individual thinks analytically about anything relating to their professional practice with the intention of gaining insight and using the lessons learned to maintain good practice or make improvements where possible’ [1]. Sandars [2] suggests three main aims to reflection – ‘reflection for learning’, ‘reflection to develop a therapeutic relationship’ and ‘reflection to develop professional practice’. These can enable practitioners and students to identify learning needs, support professional growth and improve practice through increased self-awareness, as well as highlighting opportunities to enhance quality of patient care and safety within organisations [3]. Professional bodies advocate that ‘reflection should be part of a doctor’s everyday practice’ [1], and it is required as part of doctors’ training and revalidation in education systems all over the world [4].

## Practice points

- GenAI chatbots can produce high-quality reflective writing that can evade human and automated detection.
- Health professions students and clinicians are using GenAI to generate reflections.
- GenAI may support reflective practice through scaffolding and conversational guidance.
- The validity of asynchronous reflective writing as a summative assessment is increasingly uncertain.
- In-person assessment may offer a more valid alternative.

Achievement of these aims is often assessed through asynchronous written reflections [5,6]. Charon [7] suggests that writing reflectively about patients allows for deeper understanding and self-reflection

**CONTACT** Philip M. Newton  [p.newton@swansea.ac.uk](mailto:p.newton@swansea.ac.uk)  Swansea University, Singleton Park Campus, Swansea, SA2 8PP, UK.

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/0142159X.2026.2658172>.

© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

compared to other methods, whilst Shapiro, Kasman [8] comments that the act of writing allows learners to consider a broader perspective. In addition, the act of writing reflection may lead to new ideas and insight about an event and allow for reconsideration of the facts [9].

ChatGPT (Chat Generative Pre-Trained Transformer) [10] is estimated to have 700 million weekly users since release in November 2022 [11]. Similar chatbots have been released since, including Microsoft's Copilot [12] and Google's Gemini [13]. These chatbots are powered by large language models (LLMs) which are AI systems trained on extensive amounts of data that can produce responses mimicking human language [14], hence the term 'generative' artificial intelligence (GenAI). The LLM which powers ChatGPT has been updated multiple times since its release [11], owing to significant improvements in the technology [15]. Through a simple interface, users can provide chatbots with a question or set of instructions, known as a 'prompt' [16], and the chatbot will generate a response. 'Prompt engineering' refers to the ways in which LLMs are instructed by prompts, and the interactions with, and output of the chatbots is dependent on the initial and subsequent prompts [16].

GenAI has shown great potential within medical education, from performing well on the UK Medical Licensing Assessment [17], to acting as a virtual patient for consultations [18]. However, use of this technology within education raises concerns around academic integrity [19]. Surveys examining ChatGPT use amongst medical and pharmacy students in the United States found that 48.9% [20] and 30.2% [21] had used ChatGPT for academic purposes respectively.

A review of GenAI policies in higher education assessment found that originality of students' work is viewed in relation to academic misconduct, and a spectrum exists ranging from complete prohibition to conditional permitted use [22]. Given the potential for GenAI to write reflectively, and the variation that exists in academic policy across institutions, this scoping review aims to answer the question as to whether reflective writing is a valid summative assessment in the era of GenAI. We do this by identifying papers which ask the following:

- Is reflective writing a valid summative assessment in the age of GenAI?
- Can GenAI write reflectively?
- What properties do GenAI-written reflections have?
- Can GenAI-written reflections be differentiated from those written by humans?
- Can GenAI be incorporated into the process of reflective writing for summative assessment?

## Methods

A scoping review was used to address the research question. A scoping review is designed to identify the current evidence base, clarify important concepts, ascertain key characteristics, and determine knowledge gaps, particularly in emergent concepts [23]. We use the framework originally defined by Arksey and O'malley [24], and further developed by Levac, Colquhoun [25]. The PRISMA-ScR [26] standards were followed, with references to updated information [27,28]. The five stages defined by Arksey and O'malley [24] are:

- Stage 1: identifying the research question
- Stage 2: identifying relevant studies
- Stage 3: study selection
- Stage 4: charting the data
- Stage 5: collating, summarising and reporting the results.

### Identifying the research question

The over-arching research question for this scoping review was: *'Is reflective writing a valid summative assessment in the era of generative AI?'*. There are many related research questions which can influence the answer to this over-arching question and so we identified papers which tested the following:

- Can GenAI generate reflective writing?
- Is it possible to detect reflective writing if it has been produced using ChatGPT, either by using detection software or human-based detection?
- Are health professionals or students using GenAI for reflective writing assessments?
- Can GenAI be used to support the reflective process?

The 'Population, Concept, Context' framework [27] was used to help guide the search strategy and inclusion criteria as follows:

### Inclusion criteria

- Population – learners or educators in health professions education (e.g. medical, nursing, pharmacy, dentistry, allied health professionals)
- Concept – use of GenAI chatbots, studies exploring the role of GenAI in producing or supporting reflective writing/practice including surveys of health professionals using it for this purpose
- Context – health professions education (undergraduate or postgraduate including continuous professional development) worldwide, any form of reflective writing/practice as either formative or summative exercises
- English language only
- 2022-present

PRISMA 2020 flow diagram for new systematic reviews which included searches of databases, registers and other sources

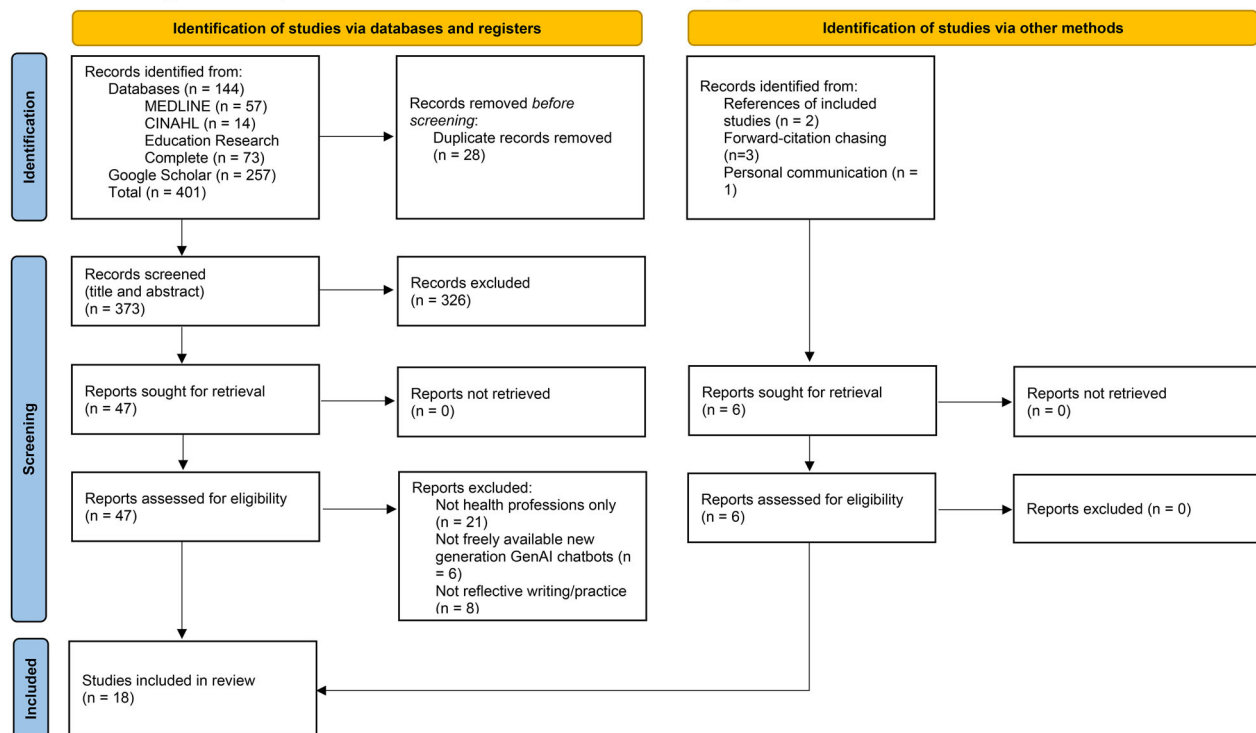


Figure 1. PRISMA flow diagram.

Exclusion criteria were the inverse of the inclusion criteria, and studies utilising tailored AI-tools were excluded.

Medline, CINAHL, Education Research Complete electronic databases and Google Scholar were searched in July 2025 using the following terms:

#### Database search terms

- 'Generative artificial intelligence' OR 'generative AI' OR genAI OR 'artificial intelligence' OR 'large language model' OR 'large language models' OR LLM OR ChatGPT OR GPT-4 OR copilot OR claude OR gemini) AND (Reflect\* N1 (writing OR write OR essays OR essay OR log OR practice OR self)'

#### Google Scholar search terms

- 'Generative AI' AND 'reflective practice' AND 'medical students' OR 'health professions'

The reference lists of all included articles were hand-searched, and forward citation chasing was conducted, whereby reviewing papers that had subsequently cited the key articles, to identify publications not included in the initial search. This process was conducted between 16th July 2025 and 12th August 2025. Studies were included if they were published in 2022 or later.

#### Study selection

Titles and abstracts were screened against eligibility criteria. If it was unclear whether a study fulfilled the

criteria then the full text was reviewed. PRISMA flow diagram is shown in Figure 1. The aim of a scoping review (compared to a full systematic review) is, as originally stated by Arksey and O'Malley, to undertake an analysis to guide a narrative account of the literature. They go on to state that 'emphasis is not placed on the 'weight of evidence' nor on evaluating the quality of evidence'. Thus, we did not limit our study selection based on quality metrics such as sampling method, blinding etc. We included all publications that met the inclusion criteria, and then considered the quality of the evidence as part of our analysis and discussion.

#### Charting the data

The following were extracted from each paper: author, year of publication, country, journal (including whether it was a predatory journal), type of publication (journal article, conference paper), title, health-professions speciality (e.g. medical students, pharmacists, dentists etc), study aims, study design (e.g. cross-sectional, systematic review, commentary etc), sampling methods (if applicable), piloting of surveys (if applicable), ethical approval (if applicable), type of GenAI (e.g. ChatGPT, Copilot, Gemini etc) and version used (if specified), date of data collection (if applicable), response rate (if applicable), sample size (if applicable), applications of GenAI (e.g. as a conversational partner, to write reflections etc), main findings, limitations of GenAI, limitations of the

study, conclusions and future directions. Data charting was done independently by one author (EF) with any uncertainty resolved through discussion with the second (PMN).

Quality appraisal was undertaken through critical reading of the studies by both authors. Given the novelty of the question and the broad range of methodologies considered, there were no obvious established quality appraisal tools that could be used. Quality features are noted within the results section where they affect the interpretation of the findings.

### **Collating, summarising and reporting the results**

A thematic analysis was conducted to identify common themes within the studies, using the six-step method of Braun and Clarke [29]. One author (EF) made extensive notes about each of the included studies in discussion with the second author (PMN). These notes were on topics relating to the research questions and were used to develop provisional themes.

The notes were then used to undertake an adjunct thematic analysis using ChatGPT running GPT-5, based on prior studies which have identified effective prompting for this purpose, and which have demonstrated that ChatGPT is effective for this task [30–32]. ChatGPT-5 Plus in ‘thinking’ mode was used as this ensures that ChatGPT ‘thinks longer for better answers’ [33]. Analysis was undertaken on 23 August 2025 and the detailed prompting strategy is in [supplementary material S1](#). In summary, the first prompt was ‘can you give me a short summary (2 paragraphs) of the contents of the uploaded file please?’, to ensure that ChatGPT was analysing the document correctly. ChatGPT was then asked to perform a thematic analysis in accordance with Braun and Clarke’s methodology [29]. This was done using three variations of prompts in three separate chats to see how this would affect ChatGPT’s output, and whether any new themes emerged. The middle section of each prompt varied, but the beginning and end of the prompt remained the same throughout.

The authors compared the provisional themes generated by EF with those generated by ChatGPT, which showed remarkable similarity. The provisional themes were merged with those generated by ChatGPT to create a final set of themes which were then discussed and agreed by both authors.

## **Results**

18 publications were included. The PRIMSA flow diagram is shown in [Figure 1](#), and a summary of the included studies is in [Table 1](#).

### **Thematic analysis**

The final set of themes is in [Table 2](#), with a mapping to identify which studies were included.

#### ***GenAI can write reflections that cannot be meaningfully distinguished from human-authored texts***

Multiple studies examined the ability of GenAI to produce reflective writing. Li, Sha [34] generated reflections using nine prompting strategies and compared these to reflections written by students prior to the launch of ChatGPT. Student reflections were based on students’ feelings of a prior learning experience, using Driscoll’s model [35]. Reflections generated by AI were slightly shorter, but were marked higher on an established rubric with an overall score of 4.3 (out of 6, SD = 0.99) for ChatGPT reflections compared with 2.9 (out of 6, SD = 1.72) for human reflections, although the sample was small ( $N=45$  for each author type) [34]. Neither human nor tech-based detection systems could reliably distinguish between human and AI generated reflections, even across a much larger sample ( $N=180$  per author type) [34]. Similar problems with detection was reported by Wraith, Carnegie [36], using Microsoft Co-Pilot but with a very small sample ( $N=4$  per author type, ratings from 28 educators). Conversely, Brondani, Alves [37] showed that three experienced instructors could identify the author as ChatGPT or human ~85% of the time, although the sample was very small ( $N=20$  per author type) and there was no way to guarantee that the human-authored reflections had not been generated with the assistance of AI and the quality of the reflections was not determined. The authors deliberately used reflections that were ‘generic’ and written in the third person. All studies appeared to use unedited outputs from GenAI and neither Li nor Brondani state the model used, although the date of the research suggests it was likely GPT-3.5 or GPT-4. Nabasenja, Chau [38] showed that ChatGPT-4o could produce reflections on clinical radiography placement experiences. The standard of the reflections aligned with professional standards in diagnostic radiography in Australia, although they lacked some properties as outlined below in section ‘Authenticity and emotional nuance of AI-generated reflections’. In all studies it was unclear whether markers were blinded to the source of the reflective writing when undertaking their analysis.

Kaliterna, Žuljević [39] et al. aimed to see whether ChatGPT and Google’s Bard could produce unstructured reflective essays on medical students’ experiences, and whether these were distinguishable from human-written texts. Fifth- and sixth-year medical students wrote essays reflecting on moral, ethical or professional dilemmas they encountered during their

**Table 1.** Summary of included studies.

Study number	First Author	Date of Publication	Country	Study design	Discipline	Level of education	Type of GenAI and version (if specified)	Key findings	Limitations	Conclusions
1	Alexander, K.M	29 April 2025	USA	Commentary	Pharmacy	Undergraduate	Microsoft Copilot (not specified)	Customised reflective writing prompts can help students craft personal reflections, identify gaps in responses & build confidence through providing immediate feedback	No empirical data, no assessment of learning outcomes, no comparative analysis	AI as a conversational partner may facilitate more detailed self-reflection and help students to focus on the reflective process rather than 'correctness'
2	Almazrou, S	16 August 2024	Saudia Arabia	Cross-sectional survey	Medicine, Nursing, Pharmacy, Physiotherapy	Undergraduate	ChatGPT (not specified)	ChatGPT used regularly by students. Nursing students rate it most useful for reflective practice. Younger students feel less useful for reflective practice. ChatGPT used regularly by faculty. Rated 3.81 (out of 5) for effectiveness at reflective practice. Lecturers rated it more useful than professors.	Limited generalisability. No assessment of longitudinal outcomes. No control group to infer causal relationships. Risk of selection bias. Limited generalisability. Risk of social desirability bias. No assessment of longer term outcomes. Lack of mixed methods or qualitative depth	Integrating ChatGPT into medical education can foster critical thinking skills and offer opportunities for students to engage in reflective practice
3	Basri, W.S	17 March 2025	Saudi Arabia	Cross-sectional survey	Medical Faculty	Continuing Professional Development	ChatGPT (not specified)	ChatGPT used regularly by faculty. Rated 3.81 (out of 5) for effectiveness at reflective practice. Lecturers rated it more useful than professors.	Limited generalisability. Risk of social desirability bias. No assessment of longer term outcomes. Lack of mixed methods or qualitative depth	Medical educators feel that ChatGPT is modestly helpful in facilitating reflective practice
4	Furey, P	09 May 2025	UK	Commentary	Paramedics	Continuing Professional Development	ChatGPT (not specified)	GenAI can be used to scaffold reflection and adjust to learner level, create group discussion prompts and simulated patient cases for discussion	Lack of empirical data and comparative analysis. No formal assessment on how GenAI impacts. No feedback from paramedics or educators	GenAI can scaffold the reflective process, streamline self-directed learning & CPD documentation
5	Khojasteh, L	31 January 2025	Iran	Instrument Development and Validation	Medicine	Undergraduate	ChatGPT (not specified)	Use of AI toolkit fostered critical reflection of writing abilities, enabled feedback to be used more effectively and increased awareness of own strengths & weaknesses	Single site. No subgroup analysis due to homogeneity. Unsure how use of the tool is affected by demographics or writing experience	Self-assessment toolkit enhances writing proficiency/reflective practice.
6	Lewis, M	23 August 2024	UK	Qualitative Case Studies	Medicine	Continuing Professional Development	ChatGPT 4	GenAI can aid in deeper reflection by asking unexpected questions, depending on prompts. Using GenAI is less time-consuming, can be more accessible, and supports reflection by encouraging deeper introspection	Small sample size. No longitudinal data. No comparison with other methods such as Balint or peer groups.	GenAI could foster deeper reflection in GP trainees and GPs by supplementing traditional reflection, saving time and resources where there may be scarcity

(continued)

Table 1. Continued.

Study number	First Author	Date of Publication	Country	Study design	Discipline	Level of education	Type of GenAI and version (if specified)	Key findings	Limitations	Conclusions
7	Pham, T.D	18 June 2025	Australia	Systematic Review	Medicine, Nursing, Dentistry, Pharmacy, Veterinary	Undergraduate	ChatGPT (not specified)	Most students use GenAI for acquisition, inquiry, practice & production. 7 papers noted use for reflection	Publication bias. English-only studies. No detail about how GenAI was being used for reflective practice	GenAI could be used for reflective practice but there are concerns about lack of human interaction and lack of feedback
8	Samuel, A.	06 February 2025	USA	Pilot proof-of-concept	Health Professions (not specified)	Postgraduate	ChatGPT 4o	ChatGPT created a more detailed and scaffolded version of a reflective practice assignment. Student reflections were of higher quality using the refined assignment	Small scale pilot study. No detail of sample size. No feedback from students	GenAI can enhance reflective experiences and promote deeper reflection through redesigning assessments
9	Stratton-Maher, D.	16 July 2025	Australia	Commentary	Nursing	Undergraduate	ChatGPT (not specified)	GenAI facilitated personalised learning pathways within the portfolio assessment, which impacted student learning by offering immediate feedback on their reflections	No mention as to how ChatGPT was integrated or whether the quality of reflections improved	GenAI can transform portfolios into interactive and dynamic learning platforms, which can benefit reflective practice
10	Wang, L.	28 May 2025	Australia	Commentary	Medicine	Continuing Professional Development	ChatGPT (not specified), Microsoft Copilot (not specified)	GenAI can write reflections. Software nor humans not good at detecting AI input	No primary data/ Single author perspective. Untested curricular recommendations. Limited generalisability.	GenAI may undermine the purposes of reflecting
11	Warrington, D.	20 December 2024	UK	Cross-sectional survey	Medicine	Continuing Professional Development	ChatGPT (not specified)	37% UK doctors surveyed used AI for writing reflections for their portfolio. Concerns about integration in healthcare	Sampling Bias (self-selection). Unable to verify respondents	UK doctors are using AI to write reflections, without guidance. Doctors want guidance and regulation on AI use
12	Wraith, C	02 July 2025	UK	Experimental	Medicine	Undergraduate	Microsoft Copilot (GPT-4)	Educators cannot reliably distinguish between GenAI and student-authored reflections	Small sample size. Cannot guarantee GenAI not used in student reflections. Unknown educator experience with GenAI	Difficulty in differentiating between GenAI and human-authored reflections. Institutions should utilise GenAI to develop students reflective skills and use other methods of assessment such as facilitated discussions

(continued)

**Table 1.** Continued.

Study number	First Author	Date of Publication	Country	Study design		Discipline	Level of education	Type of GenAI and version (if specified)		Key findings	Limitations	Conclusions
				One-group	post-test only			ChatGPT (not specified)	(if specified)			
13	Huang, Y.Y	15 July 2025	Taiwan	One-group	post-test only	Nursing	Undergraduate	ChatGPT (not specified)		GenAI-assisted reflection can enhance nursing students' cognitive, emotional, and behavioural learning. Nursing students reported positive perceptions of this use.	No control group. No objective measures of reflection. No mention of AI prompt design or content.	GenAI tools may help structure and guide reflective practice in community health nursing education
14	Li, Y.	12 May 2023	Australia	Experimental		Pharmacy	Undergraduate	ChatGPT (not specified)		ChatGPT-generated reflections scored higher than student-written ones across all assessment criteria. 38% higher detection of GenAI-written reflections when trained software used versus human detection	Single discipline. Artificially optimised prompting strategies does not reflect student use. Lack of inter-rater reliability	ChatGPT can produce reflective responses that outperform students work. Trained software can effectively detect AI-generated content
15	Brondani, M.	07 July 2024	Canada	Experimental		Dentistry	Undergraduate	ChatGPT (not specified)		University instructors can correctly identify authorship of reflections in 85% of cases. ChatGPT reflections can be mistaken for students	Instructor familiarity with ChatGPT unknown. Unknown whether students had used GenAI in their essays.	University instructors could mostly differentiate ChatGPT-authored from student-authored reflections. ChatGPT reflections can be mistaken for students.
16	Kalitema, M	30 October 2024	Croatia	Cross-sectional survey		Medicine	Undergraduate	ChatGPT 3.5 and Bard		GenAI-written reflections are linguistically different from student-written. 34% of student essays were partially/fully written by GenAI	Reflections written by non-native English speakers could influence linguistic comparison. Ambiguity in comparing student versus GenAI essays when GenAI input into student essays is unknown. Uncertainty regarding false positives/negatives when using AI detection tools	ChatGPT and Bard capable of generating essays that resemble human-like text. GenAI essays used more authentic and affective language. Detection tools insufficient if students modify AI output
17	Nabasenja, C	07 August 2025	Australia	Experimental		Diagnostic Radiography	Undergraduate	ChatGPT-4o		GenAI reflections aligned with professional standards in radiography. GenAI reflections scored high	Small sample size of scenarios and raters. Lack of student input means unsure how	ChatGPT-4o can be a useful scaffolding tool to help students structure reflections in medical radiation

(continued)

Table 1. Continued.

Study number	First Author	Date of Publication	Country	Study design	Discipline	Level of education	Type of GenAI and version (if specified)	Key findings	Limitations	Conclusions
18	Kawakami, C.	28 April 2024	Japan	Cross-sectional survey	Medicine	Undergraduate	ChatGPT (not specified)	for clarity & structure but lacked emotional depth, person-centred insight and contextual awareness Medical students' positive perceptions of and preference for using GenAI in reflective reports increased significantly between 2023 and 2024.	students might engage with GenAI in practice Limited generalisability. Effectiveness of incorporating GenAI into reflective assessments not evaluated	science education. GenAI should complement and not replace human-guided reflections The findings highlight a rapid cultural shift amongst medical students towards the acceptance of using GenAI for reflection

studies. The student-written essays were collected in 2023, and keywords extracted from these essays were then used to develop prompts for generating GenAI reflections. Attempted detection of AI-generated essays was undertaken using 'Originality.AI' and 'GPTZero' software. Analysis ( $N = 47$  per author type) revealed that 34% of student essays were highlighted as being either partially or fully written by GenAI. A sub-analysis of these essays found that their linguistic properties were more similar to the GenAI essays than the 'true' student ones. The authors presume that these students had used GenAI in their essays.

A viewpoint published by Wang [40] contends that written reflective essays in physician training are particularly vulnerable to GenAI due to the ability of LLMs to write high-quality responses. They discuss the inaccuracies of AI-detection software to distinguish between human and GenAI-written content, which becomes even more difficult when AI-generated text is altered [41].

In summary, GenAI can generate reflective writing to a high academic standard, although they may be qualitatively different to human generated reflections as described in section 'Authenticity and emotional nuance of AI-generated reflections'. Assessors cannot reliably distinguish between student-authored and GenAI-authored reflections.

### *Authenticity and emotional nuance of AI-generated reflections*

At least two opinion pieces argue that GenAI cannot simulate the lived experience, nuance and complex cognition required for reflective practice (Furey [42], Lewis and Hayhoe [43]). However when comparing linguistic properties of ChatGPT and Bard-generated essays to medical student essays, Kaliterna, Žuljević [39] et al. found that AI-generated essays contained significantly more 'affect' related words, particularly relating to positive emotion. Bard's text scored significantly higher on the 'social processes' variable and was more human-like. In comparison, ChatGPT's essays seemed more authentic due to the significantly higher use of more complex words and analytical thinking. Student-written essays contained significantly more words-per-sentence and used more cognitive language.

However, Nabasenja, Chau [38] found that reflections created by ChatGPT-4o scored highest for 'clarity & structure' but lowest for 'depth & insightfulness'. Qualitative comments provided by the markers revealed that the reflections were lacking empathy, not person-centred, superficial and lacking depth, nuance and complexity. However, the very small sample size, confidence intervals and unblinded

**Table 2.** Themes including descriptions and summary of key findings.

Theme	Description	Summary of key findings	Study numbers
GenAI Can Write Reflections That Cannot Be Meaningfully Distinguished from Human-Authored Texts	Explores the ability of LLMs to write reflections and whether humans or computational classifiers can distinguish between the reflective writing from humans vs GenAI	GenAI-authored reflections were rated higher than human-written reflections and could pass undetected by humans and software	10, 12, 14, 15, 16, 17
Authenticity and Emotional Nuance of AI-generated Reflections	Explores the authenticity of generative AI-authored reflections with a focus on their emotional and linguistic characteristics	GenAI-authored reflections tended to use affective language but were perceived as lacking emotional depth when compared to human-written reflections	4, 6, 12, 14, 15, 16, 17
Adoption and Patterns of Use	Explores the prevalence and purpose of generative AI use amongst health professions education	Studies highlighted that health professionals, students and faculty have frequently used GenAI, with a proportion having used it to write reflections for their portfolio	2, 3, 6, 7, 11, 18
Ethics, Governance and Responsible Use	Explores ethical concerns around generative AI use and the desire for increased guidance	Studies highlight risks related to privacy with regards to patient confidentiality, as well as concerns around academic integrity and transparency of use amongst students and staff	2, 3, 4, 5, 7, 8, 10, 11
Generative AI as a Reflective Scaffold and Coach	Explores how generative AI can be used to support reflective practice	GenAI can support reflection by acting as a conversational partner by offering structured prompts and encouraging deeper questioning	1, 2, 3, 4, 5, 6, 7, 8, 9, 13
Redesigning Reflective Assessments	Explores how reflective assessments can be redesigned to ensure validity is maintained	Studies outlined ways in which reflective assessments could be adapted to either incorporate GenAI or reduce the risk of use	4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15, 16, 17, 18

nature of the analysis may limit the generalisability of these findings.

Wraith, Carnegie [36] interviewed markers of GenAI-authored reflections who reported that GenAI essays contained perfect spelling and grammar, whereas student essays contained mistakes, were poorly constructed and lacking in flow. However, in 'features of reflection', student reflections discussed emotions and identified aims for improvement, compared to GenAI reflections that lacked detail, authenticity and were too articulate.

In summary, GenAI can produce polished reflections that score highly on emotional criteria, and there is concern that they are superficial by lacking the lived experience and depth that is crucial to reflecting.

### **Adoption and patterns of use**

Warrington and Holm [44] surveyed 211 UK doctors in 2024 and found more than 50% of doctors had used AI, with younger doctors and doctors-in-training more likely to have used it. Strikingly, the most common use was for writing reflective pieces for their assessment and continuing professional development portfolios, with 37.1% of GenAI users deploying it for this purpose. Other survey studies show that both students [45] and faculty Basri, Attar [46] are using GenAI very frequently, but participants showed modest levels of agreement when asked whether GenAI can be used effectively for reflective

practice. These survey studies use convenience samples with unvalidated instruments and so it is unclear how representative the results are.

A systematic review by Pham, Karunaratne [47] assessed the impact of GenAI on health professions education and mapped this against Laurillard's Conversational Framework [48]. They found that 73% of studies used GenAI to 'practice' something, such as virtual patient consultations, communication skills, managing medications as well as reflective practice. Other use cases included writing essays, simplifying complex information and correcting errors.

Thus, it appears that health professionals and students use GenAI, including to generate reflective writing.

### **Ethics, governance and responsible use**

Concerns around the ethics and governance of GenAI use was a prominent theme in many studies, with surveys and commentaries commonly reporting high levels of concern about patient confidentiality, data privacy, output accuracy and overreliance on GenAI (Almazrou, Alanezi [45], Basri, Attar [46], Pham, Karunaratne [47], Warrington and Holm [44]). Authors called for professional bodies to issue guidance around GenAI use, particularly around these ethical issues. For example, the 'AI Assessment Scale' which details how GenAI can be used in assessment and appears to reduce misconduct (Furze et al. 2024 as cited in Wang [40]). However, doctors who

reported greater knowledge of AI were less supportive of AI regulation (Warrington and Holm [44]).

Khojasteh, Kafipour [49] advocates 'ethical and responsible AI integration', where students should be taught how to critique GenAI's output and should use it to supplement, and not substitute, their own learning. Concerns regarding biases are also echoed, as these may be unintentionally perpetuated by GenAI algorithms based off their training data [49].

Honesty around using GenAI is important not only for students, but for educators too. Samuel, Soh [50] used ChatGPT to redesign reflective practice assessments. They discuss the importance of being transparent with learners when GenAI has been used to design their assessments, thus facilitating discussions about ethical use.

### *Generative AI as a reflective scaffold and coach*

The use of GenAI to support reflection through providing tailored support or acting as a coach was a theme present in over half of the studies included [42,43,45–47,49–53], although evidence to support this role was limited. Both Alexander, Johnson [51] and Furey [42] used case study approaches, for example using Socratic prompting, tailored to specific clinical experiences and expertise level. They also suggest some potential benefits such as instant feedback, and allowing the learner to focus on the process, rather than the grade.

Some studies provided empirical evidence to support these ideas [43,47,49,50,52,53]. Lewis and Hayhoe [43] presented three different case studies to show how ChatGPT can 'formulate insightful and challenging questions' for different levels of experience within GP. Using voice input technology, case vignettes (based on real patient encounters) were narrated into ChatGPT-4, aimed at different levels of professional experience. One vignette asked ChatGPT to guide the user through Gibbs Reflective Cycle [54] and to encourage deeper reflection by asking two probing questions. Another was designed for a more experienced practitioner reflecting on a difficult case and asks ChatGPT to integrate Brian and Eno's 'Oblique Strategies' (prompts to inspire creative thinking amongst artists). This approach generated unexpected questions which prompted reflection from new perspectives. Another vignette used ChatGPT's real time voice abilities, where the GP dictated their experience into ChatGPT, employing an interactive conversational approach. The authors recommended that GenAI could guide individuals towards more personal, authentic reflection by encouraging deeper introspection. However, there is no data comparing this approach to traditional reflective methods or exploring whether using GenAI

for reflection would be considered acceptable by practitioners or assessors.

Pham, Karunaratne [47]'s systematic review examined the impact of GenAI on health professions education (HPE). Studies were included if they specifically mentioned how HPE students engaged with GenAI, and quality was assessed using the Medical Education Research Study Quality Instrument. 73% of studies described how GenAI could be used for 'practice', which was defined as 'actions that were repeated and accompanied with feedback cycles'. Within this domain, six studies noted students using GenAI for reflective practice. However, no further detail is provided as to how GenAI was used for reflection within those studies, and the authors do not define reflective practice. The cited studies do not all appear to relate to reflective practice as described here, and those that do are included here.

Studies by Stratton-Maher [52] and Huang, Lin [53] reviewed how students could use ChatGPT to help write their reflections. Stratton-Maher [52] reviewed the impact of integrating ChatGPT into portfolio assessment for the 'literacy and communication for healthcare' course within a Nursing program. Students were instructed how to effectively use ChatGPT within their assessment and were able to use it to structure their language and generate reflections for their portfolios. Stratton-Maher [52] suggests that this approach could transform portfolios into interactive platforms to facilitate personalised learning, although the quality of the reflections was unclear. Similarly, Huang, Lin [53] evaluated ChatGPT as a tool to support reflective practice at the end of a course on community health nursing. Students used ChatGPT to guide them through six reflective stages to create accounts based on their caregiving experiences before completing a self-reflection questionnaire about AI-assisted reflection. Both Stratton-Maher [52] and Huang, Lin [53] highlighted the ability of ChatGPT to provide instant feedback and personalised writing assistance. Huang, Lin [53] also described how ChatGPT asked reflective questions and generated summaries, with students reporting good engagement, although there was no control group, no pre-post comparison and again the quality of the reflections was unclear.

Khojasteh, Kafipour [49] and Samuel, Soh [50] suggest that GenAI can scaffold reflection and improve the quality of students' work through improving structure and self-assessment. Khojasteh, Kafipour [49] asked medical students to write short essays on various topics. Students then prompted GenAI to 'review and revise' their essay, suggesting and justifying improvements to coherence, punctuation, grammar and vocabulary. Medical students then

wrote reflections about their experiences of using GenAI. These essays underwent thematic analysis, which was then discussed in expert focus groups as part of the toolkit creation process. A comparative evaluation was conducted of student's self-assessment scores related to the GenAI feedback both with and without the AI self-assessment toolkit. The results showed that the toolkit demonstrated high reliability, face and content validity. Use of the toolkit supported the reflective process by improving students' ability to critically assess their own writing, set achievable goals and utilise feedback more effectively. Feedback was also aligned with teacher feedback, allowing for more accurate self-assessment. Samuel, Soh [50] showed how GenAI can scaffold reflection by using ChatGPT-4o to redesign reflective practice assessments in graduate-level health professions education courses. Using ChatGPT to redesign the reflective assignment provided more detail and scaffolding for students than previous versions, as noted by the course instructor. Student assignments produced before and after the assessment redesign were assessed by two independent faculty members that were blinded to the study and revealed that the quality of the students' work had improved, although outcomes were not quantified and sample size not reported.

### **Redesigning reflective assessments**

Many studies advocated changing the way in which the reflection is assessed [34,36–40,42–44,47,49,50,52,55], while allowing learners to use (or not) GenAI during the reflective process. For example Warrington and Holm [44] suggest it could be acceptable for GenAI to write reflections that doctors then reflect upon, as long as this is acknowledged, and the clinicians discuss their reflections with colleagues [The Schwartz Centre. Schwartz rounds, as cited in 44]. A similar argument is made by both Wang [40] and Wraith, Carnegy [36] who suggest that the traditional reflective essay could be replaced with discussion-based activities, acknowledging that GenAI can be a helpful tool for those who find writing reflections difficult. This means of assessment would also have the added benefit of removing the expectation on faculty members to identify GenAI authored reflections and instead focus more on appraising the quality of the reflections. Similar conclusions were reached by Nabasenja, Chau [38], who suggest that students could engage in the reflective process by editing reflections written by GenAI, or by analysing human versus GenAI-written reflections. Brondani, Alves [37] also propose a human-GenAI collaborative approach, suggesting that GenAI could be used as a 'springboard' when beginning an assignment. The authors suggest that

students could prompt GenAI to write the initial reflection, before editing and personalising its output to add their own perspectives. However, it was unclear how these human-GenAI collaborative reflections could be meaningfully assessed.

Li, Sha [34] and Kaliterna, Žuljević [39] both advocate the refining of detection software, based in part on their findings that the linguistic properties of GenAI reflections are different, and thus novel detectors could be trained on these findings. Kaliterna, Žuljević [39] advocate for the return of hand-written essays as a credible alternative, to ensure students engage with the reflective process.

Many studies build on the earlier content about using GenAI as a scaffold or reflective coach, providing recommendations for how this could be developed into the assessment of reflection. For example Furey [42] advocated for research into profession-specific models that integrate aspects of clinical work such as language, professional standards and ethical subtlety, while Kawakami, Nomura [55] suggest using GenAI for more creative purposes such as creating images. However further work is needed to understand how these novel assessments could be assessed.

Khojasteh, Kafipour [49], writing about assignments generally, suggest that educators could use GenAI feedback to supplement human feedback, where GenAI could highlight technical issues such as grammar, allowing human markers to focus on content. Students could use GenAI tools in their assignments before reflecting on its output, to encourage critical thinking and reduce the risk of over-reliance. Lewis and Hayhoe [43] make similar suggestions, based on a view that integration of GenAI within medical practice is inevitable, and that we should exploit its potential as opposed to seeing it as a threat. Samuel, Soh [50] also argued for GenAI as a supplemental marker to save faculty workload and allow human markers to concentrate on content. They identify some practical steps for doing this, but more rigorous testing is needed, with control groups and quantitative outcomes. Similarly Stratton-Maher [52] state that incorporating ChatGPT into portfolio assessments 'significantly impacted student learning' and deepened students 'understanding of nursing concepts', but don't give further details or objective outcomes as to how GenAI did this.

## **Discussion**

Reviewed studies showed that GenAI is able to generate high quality reflective writing that is difficult to meaningfully distinguish from human-authored reflective writing [34,36–40]. Clinicians and students are using GenAI for reflective writing. Thus, it is

difficult to justify the continued use of asynchronous written reflections as a summative assessment. Studies also suggested how the situation might be addressed, in the context of the literature, such as refining detection software [34]. However, the rapid evolution of the technology means that this seems unlikely to be effective and the idea seems to have fallen from favour. Hand-writing essays [39] may also not be a viable solution since students could still use GenAI where they hand-write their essays.

Rather than try and ban, or detect GenAI use in written reflections, many studies advocated for embracing GenAI, bring it into the process and assessment of reflection [36–38,42–44,49–53,55]. However, the evidence that GenAI is effective was limited, based mostly on expert opinion and qualitative case studies. Rigorously designed experimental studies testing the efficacy and safety of GenAI as a reflective partner would give educators more confidence (or not) in this approach. Many studies articulate both policy and ethical concerns that can impact this approach, and the conflict between these embracing GenAI and these ethical issues was not really addressed in any meaningful detail.

To address this conflict, authors call for clear guidance about the use of GenAI within assessments and portfolios, but, at the time of writing, it is far from clear what that guidance might look like. For example the International Association for Health Professions Education published a guide about AI, aiming to highlight how educators can embrace its potential in health professions education assessment [56]. They acknowledge that the ‘take-home essay’ is particularly vulnerable to AI input but suggest that banning or minimising the use of GenAI does not equip learners with the skills needed to navigate the real world and risks losing sight of some of the key benefits that AI can have. Instead, Masters, MacNeil [56] argue that we should embrace AI and consider other methods of assessing learning. One example is to see the written essay as one component of the assessment process, rather than the only means of assessment. Learners could utilise AI to create text, images and audio to meet their assignment brief. However, whether this is feasible with regards to reflective assessments remains to be seen.

A study by Neshaei, Mejia-Domenzain [57] showed how GenAI (GPT-4o and a fine-tuned BERT model) can be used to support reflection. They built an AI-enabled reflective writing assistant called ‘Reflectium’, a web-based application which guided health and welfare learners through Gibbs Reflective Cycle, allowed for practice and gave sentence-by-sentence feedback. They concluded that using customised GenAI tools can support learning through development of metacognitive skills, and assist

learners to write more structured, reflective texts. Whilst this is only one study, there is potential for scalability and utility in health professions education across the board, and future research could explore these customised tools unlike the freely available ones reviewed here.

Another approach might be to adapt the way in which reflection is assessed. As suggested by some of the articles within this review, one answer could be to introduce assessed discussions, either with a supervisor or in a group setting. Discussion could prompt learners to reflect more deeply on their experiences through thinking aloud. Even if the learner uses GenAI to create their reflection, discussing this with another would require them to reflect on GenAI’s output. However, this approach could be resource-intensive.

### Limitations

There are multiple limitations to the review, and our conclusions. Perhaps most striking is that multiple studies did not specify the version of ChatGPT used, or the date in which it was used. This is particularly important as the LLMs powering ChatGPT now are more sophisticated than those used prior [11,15]. Most studies also used modest samples, and it was unclear whether markers were blinded when grading assignments. Sampling bias through convenience sampling may have affected results, and social desirability bias was often not acknowledged. Small sample sizes may affect generalisability of findings. Lack of comparison with control groups limits the ability to infer causal relationships and control for confounding variables. We also included expert opinion commentary-type articles which did not have empirical data but given the speed and scale with which GenAI is developing, and the slow pace of peer review making primary research out of date quite quickly, the authors considered it important to include expert opinion. A set of protocols for educational benchmarking would greatly aid the field [58].

### Conclusion

It appears difficult to justify asynchronous reflective writing as a valid summative assessment. GenAI chatbots such as ChatGPT can write reflections to a high academic standard that can evade detection by humans and software, and it is being used for this purpose by health professions students and clinicians. This undermines any learning that comes from reflection, and the means that learning cannot be certified in any reliable or valid way using written reflection. In-person assessment of reflective writing might be a viable alternative in the short term, while

future research could robustly test whether GenAI can meaningfully support the reflective process while protecting learning and respecting ethical and privacy concerns.

## Acknowledgements

The authors would like to thank the subject librarians at Swansea University for their expert support in refining the search strategy for the literature review. Both authors contributed to study conception and design. Supervision was provided by PMN throughout. The literature search and data analysis was performed by EF. The first draft of the manuscript was written by EF, with review and editing performed by PMN. Both authors read and approved the final manuscript.

## Authors contributions

CRedit: **Emma Furzer**: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Visualization, Writing – original draft; **Philip M. Newton**: Conceptualization, Methodology, Project administration, Supervision, Validation, Writing – review & editing.

## Disclosure statement

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

## Funding

No funding was received for this study.

## Notes on contributors

**Emma Furzer**, MBBCh, is a postgraduate student in Medical Education at Swansea University.

**Philip Newton**, is Professor of Medical Education and Neuroscience at Swansea University, with interests in neuroscience, evidence-based education and artificial intelligence within education.

## ORCID

Emma Furzer  <http://orcid.org/0009-0002-4521-5569>

Philip M. Newton  <http://orcid.org/0000-0002-5272-7979>

## References

- Academy and COPMeD Reflective Practice Toolkit C. o. P. M. D. Academy of Medical Royal Colleges 2018 DOI: [https://www.aomrc.org.uk/wp-content/uploads/2018/09/Reflective\\_Practice\\_Toolkit\\_AoMRC\\_CoPMED\\_0818.pdf](https://www.aomrc.org.uk/wp-content/uploads/2018/09/Reflective_Practice_Toolkit_AoMRC_CoPMED_0818.pdf)
- Sandars J. The use of reflection in medical education: AMEE Guide No. 44. *Med Teach*. 2009;31(8):685–695. doi: [10.1080/01421590903050374](https://doi.org/10.1080/01421590903050374)
- General Medical Council. The reflective practitioner guidance for doctors and medical students; 2020.
- Lim JY, Ong SYK, Ng CYH, et al. A systematic scoping review of reflective writing in medical education. *BMC Med Educ*. 2023;23(1):12. doi: [10.1186/s12909-022-03924-4](https://doi.org/10.1186/s12909-022-03924-4)
- Chretien KC, Chheda SG, Torre D, et al. Reflective writing in the internal medicine clerkship: a national survey of clerkship directors in internal medicine. *Teach Learn Med*. 2012;24(1):42–48. doi: [10.1080/10401334.2012.641486](https://doi.org/10.1080/10401334.2012.641486)
- Charon R, Hermann N. Commentary: a sense of story, or why teach reflective writing? *Acad Med*. 2012; 87(1):5–7. doi: [10.1097/ACM.0b013e31823a59c7](https://doi.org/10.1097/ACM.0b013e31823a59c7)
- Charon R. Narrative medicine: honoring the stories of illness. Oxford: Oxford University Press; 2008.
- Shapiro J, Kasman D, Shafer A. Words and wards: A model of reflective writing and its uses in medical education. *J Med Humanit*. 2006;27(4):231–244. doi: [10.1007/s10912-006-9020-y](https://doi.org/10.1007/s10912-006-9020-y)
- Launer J. What's the point of reflective writing? *Postgrad Med J*. 2015;91(1076):357–358. doi: [10.1136/postgradmedj-2015-133514](https://doi.org/10.1136/postgradmedj-2015-133514)
- OpenAI. Introducing ChatGPT: openAI. 2022. <https://openai.com/index/chatgpt/>.
- Chatterji A, Cunningham T, Deming D, et al. How people use ChatGPT. Cambridge, MA: National Bureau of Economic Research; 2025.
- Microsoft. Microsoft Copilot [AI assistant]. Redmond, WA: Microsoft; 2023.
- Google. Gemini [large language model]. Mountain View, CA: Google DeepMind; 2023.
- Brown T, Mann B, Ryder N, et al. Language models are few-shot learners. *Adv Neural Inf Process Syst*. 2020;33:1877–1901.
- Minaee S, Mikolov T, Nikzad N, et al. Large language models: A survey. *arXiv preprint arXiv:240206196*; 2024.
- White J, Fu Q, Hays S, et al. A prompt pattern catalog to enhance prompt engineering with chatgpt. *arXiv preprint arXiv:230211382*; 2023.
- Lai UH, Wu KS, Hsu T-Y, et al. Evaluating the performance of ChatGPT-4 on the United Kingdom medical licensing assessment. *Front Med*. 2023;10: 1240915. doi: [10.3389/fmed.2023.1240915](https://doi.org/10.3389/fmed.2023.1240915)
- Aster A, Ragaller SV, Raupach T, et al. ChatGPT as a virtual patient: written empathic expressions during medical history taking. *Med Sci Educ*. 2025;35(3): 1513–1522. doi: [10.1007/s40670-025-02342-7](https://doi.org/10.1007/s40670-025-02342-7)
- 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](https://doi.org/10.1080/0142159X.2024.2314198)
- Zhang JS, Yoon C, Williams DKA, et al. Exploring the usage of ChatGPT among medical students in the United States. *J Med Educ Curric Dev*. 2024;11: 23821205241264695. doi: [10.1177/23821205241264695](https://doi.org/10.1177/23821205241264695)
- Anderson HD, Kwon S, Linnebur LA, et al. Pharmacy student use of ChatGPT: a survey of students at a U.S. School of Pharmacy. *Curr Pharm Teach Learn*. 2024;16(11):102156. doi: [10.1016/j.cptl.2024.102156](https://doi.org/10.1016/j.cptl.2024.102156)
- Luo J. A critical review of GenAI policies in higher education assessment: A call to reconsider the “originality” of students’ work. *Assess Eval High Educ*. 2024;49(5):651–664.
- Munn Z, Peters MDJ, Stern C, et al. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med Res Methodol*. 2018;18(1):143. doi: [10.1186/s12874-018-0611-x](https://doi.org/10.1186/s12874-018-0611-x)

24. Arksey H, O'malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol*. 2005;8(1):19–32. doi: [10.1080/1364557032000119616](https://doi.org/10.1080/1364557032000119616)
25. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. *Implement Sci*. 2010;5(1):69. doi: [10.1186/1748-5908-5-69](https://doi.org/10.1186/1748-5908-5-69)
26. Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med*. 2018;169(7):467–473. doi: [10.7326/M18-0850](https://doi.org/10.7326/M18-0850)
27. Peters MDJ, Godfrey C, Mclnerney P, et al. Best practice guidance and reporting items for the development of scoping review protocols. *JBI Evid Synth*. 2022;20(4):953–968. doi: [10.11124/JBIES-21-00242](https://doi.org/10.11124/JBIES-21-00242)
28. Peters MDJ, Marnie C, Tricco AC, et al. Updated methodological guidance for the conduct of scoping reviews. *JBI Evid Synth*. 2020;18(10):2119–2126. doi: [10.11124/JBIES-20-00167](https://doi.org/10.11124/JBIES-20-00167)
29. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol*. 2006;3(2):77–101. doi: [10.1191/1478088706qp0630a](https://doi.org/10.1191/1478088706qp0630a)
30. Naeem M, Smith T, Thomas L. Thematic analysis and artificial intelligence: a step-by-step process for using ChatGPT in thematic analysis. *Int J Qual Methods*. 2025;24:16094069251333886. doi: [10.1177/16094069251333886](https://doi.org/10.1177/16094069251333886)
31. Lee VV, van der Lubbe SCC, Goh LH, et al. Harnessing ChatGPT for thematic analysis: are we ready? *J Med Internet Res*. 2024;26: E54974. doi: [10.2196/54974](https://doi.org/10.2196/54974)
32. Zhang H, Wu C, Xie J, et al. Redefining qualitative analysis in the AI era: utilizing ChatGPT for efficient thematic analysis. *arXiv preprint arXiv:230910771*; 2023.
33. OpenAI. ChatGPT [large language model]. GPT-5. San Francisco, CA: OpenAI; 2025.
34. Li Y, Sha L, Yan L, et al. Can large language models write reflectively. *Comput Educ Artif Intell*. 2023;4: 100140. doi: [10.1016/j.caeai.2023.100140](https://doi.org/10.1016/j.caeai.2023.100140)
35. Driscoll J. Practising clinical supervision: a reflective approach for healthcare professionals. New York: Elsevier Health Sciences; 2006.
36. Wraith C, Carnegie A, Brown C, et al. Can educators distinguish between medical student and generative AI-authored reflections? *Med Educ*. 2025;59(11): 1214–1221. doi: [10.1111/medu.15750](https://doi.org/10.1111/medu.15750)
37. Brondani M, Alves C, Ribeiro C, et al. Artificial intelligence, ChatGPT, and dental education: implications for reflective assignments and qualitative research. *J Dent Educ*. 2024;88(12):1671–1680. doi: [10.1002/jdd.13663](https://doi.org/10.1002/jdd.13663)
38. Nabasenja C, Chau M, Green E. Exploring ChatGPT-4o-generated reflections: alignment with professional standards in diagnostic radiography—A pilot experiment. *J Med Imaging Radiat Sci*. 2025;56(6):102082. doi: [10.1016/j.jmir.2025.102082](https://doi.org/10.1016/j.jmir.2025.102082)
39. Kaliterna M, Žuljević MF, Ursić L, et al. Testing the capacity of Bard and ChatGPT for writing essays on ethical dilemmas: a cross-sectional study. *Sci Rep*. 2024;14(1):26046. doi: [10.1038/s41598-024-77576-3](https://doi.org/10.1038/s41598-024-77576-3)
40. Wang LW. Impact of generative artificial intelligence on the professional qualities curriculum and reflection essays during physician training. *Intern Med J*. 2025;55(7):1208–1211. doi: [10.1111/imj.70087](https://doi.org/10.1111/imj.70087)
41. Perkins M. Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond. *JUTLP*. 2023;20(2):1–24. doi: [10.53761/1.20.02.07](https://doi.org/10.53761/1.20.02.07)
42. Furey P. Generative artificial intelligence 5: reflective practice. *Journal of Paramedic Practice*. 2025;17(5):1–7. doi: [10.12968/jpar.2025.0040](https://doi.org/10.12968/jpar.2025.0040)
43. Lewis M, Hayhoe B. The digital Balint: using AI in reflective practice. *Educ Prim Care*. 2024;35(6):198–202. doi: [10.1080/14739879.2024.2372606](https://doi.org/10.1080/14739879.2024.2372606)
44. Warrington DJ, Holm S. Healthcare ethics and artificial intelligence: a UK doctor survey. *BMJ Open*. 2024;14(12): E 089090. doi: [10.1136/bmjopen-2024-089090](https://doi.org/10.1136/bmjopen-2024-089090)
45. Almazrou S, Alanezi F, Almutairi SA, et al. Enhancing medical students critical thinking skills through ChatGPT: an empirical study with medical students. *Nutr Health*. 2025;31(3):1023–1033. doi: [10.1177/02601060241273627](https://doi.org/10.1177/02601060241273627)
46. Basri WS, Attar RW, Albagmi S, et al. Effectiveness of ChatGPT for educators professional development: an empirical study with medical faculty. *Nutr Health*. 2025;31(4):1533–1543. doi: [10.1177/02601060251315795](https://doi.org/10.1177/02601060251315795)
47. Pham TD, Karunaratne N, Exintaris B, et al. The impact of generative AI on health professional education: a systematic review in the context of student learning. *Med Educ*. 2025;59(12):1280–1289. doi: [10.1111/medu.15746](https://doi.org/10.1111/medu.15746)
48. Laurillard D. Teaching as a design science: building pedagogical patterns for learning and technology. London: Routledge; 2013.
49. Khojasteh L, Kafipour R, Pakdel F, et al. Empowering medical students with AI writing co-pilots: design and validation of AI self-assessment toolkit. *BMC Med Educ*. 2025;25(1):159. doi: [10.1186/s12909-025-06753-3](https://doi.org/10.1186/s12909-025-06753-3)
50. Samuel A, Soh M, Jung E. Enhancing reflective practice with ChatGPT: A new approach to assignment design. *Med Teach*. 2025;47(9):1438–1440. doi: [10.1080/0142159X.2025.2461534](https://doi.org/10.1080/0142159X.2025.2461534)
51. Alexander KM, Johnson M, Farland MZ, et al. Exploring generative artificial intelligence to enhance reflective writing in pharmacy education. *Am J Pharm Educ*. 2025;89(6):101416. doi: [10.1016/j.ajpe.2025.101416](https://doi.org/10.1016/j.ajpe.2025.101416)
52. Stratton-Maher D. The impact of generative AI on portfolio assessment in the bachelor of nursing program at the university of Southern Queensland. *E-Portfolios Australia*. 2024:17.
53. Huang Y-Y, Lin C-J, Chen C-M. Supporting reflective practice in community health nursing education: integrating ChatGPT as a tool. *International Conference on Innovative Technologies and Learning*; 2025. doi: [10.1007/978-3-031-98197-5\\_25](https://doi.org/10.1007/978-3-031-98197-5_25)
54. Gibbs G. Learning by doing: a guide to teaching and learning methods. Oxford: Further Education Unit; 1988.
55. Kawakami C, Nomura O, Takahashi M, et al. Rapidly rising preference among new-entry medical students for using generative artificial intelligence in reflective reports. *JMA J*. 2025;8(2):465–469. doi: [10.31662/jmaj.2024-0390](https://doi.org/10.31662/jmaj.2024-0390)
56. Masters K, MacNeil H, Benjamin J, et al. Artificial intelligence in health professions education assessment: AMEE guide no. 178. *Med Teach*. 2025;47(9): 1410–1424. doi: [10.1080/0142159X.2024.2445037](https://doi.org/10.1080/0142159X.2024.2445037)
57. Neshaei SP, Mejia-Domenzain P, Davis RL, et al. Metacognition meets AI: empowering reflective writing with large language models. *Brit J Educ Tech*. 2025;56(5):1864–1896. doi: [10.1111/bjet.13601](https://doi.org/10.1111/bjet.13601)
58. Newton PM, Foltýnek T. A pragmatic protocol for stress testing assessment formats against GenAI; 2024. <https://www.academicintegrity.eu/materials/393>.