



Artificial intelligence in dental education: applications, evidence and curricular challenges

Inteligencia artificial en la educación odontológica: aplicaciones, evidencia y desafíos curriculares

Maylon Felix Guerra Reyes¹  

Ernesto Learsys Campello Trujillo¹  

¹ Universidad de Ciencias Médicas de la Habana, La Habana, Cuba.

Submitted: 01-06-2025 | Revised: 03-08-2025 | Accepted: 21-10-2025 | Published: 25-10-2025

Corresponding Author: Ernesto Learsys Campello Trujillo 

ABSTRACT

Artificial intelligence is entering dental education through simulation, image-based tutoring, and, most recently, large language models, but the evidence for its educational effect and the framework for its curricular integration have not been synthesized. This narrative review traces the evolution of artificial intelligence in dental education in four stages: simulation and e-learning, image-based artificial intelligence in radiology teaching, large language models and generative artificial intelligence, and assessment integrity and curricular integration. The review draws on sources published up to 2025 and on the broader medical education literature. The evidence shows that artificial intelligence can support learning and provide feedback, but that the studies are small, heterogeneous, and mostly short-term, and that the validity of large language models for assessment is uneven. The main challenges are curricular: deciding what to teach, how to teach it, and how to preserve assessment integrity in the presence of generative tools. The review concludes that artificial intelligence should be integrated into dental curricula as a tool to be used critically rather than either embraced uncritically or excluded.

Keywords: artificial intelligence; education dental; curriculum; large language models; educational technology; students dental.

RESUMEN

La inteligencia artificial está entrando en la educación odontológica a través de la simulación, la tutoría basada en imágenes y, más recientemente, los modelos de lenguaje, pero la evidencia de su efecto educativo y el marco para su integración curricular no se han sintetizado. Esta revisión narrativa traza la evolución de la inteligencia artificial en la educación odontológica en cuatro etapas: simulación y aprendizaje en línea, inteligencia artificial basada en imágenes en la enseñanza de radiología, modelos de lenguaje y generación de contenido, e integridad de la evaluación e integración curricular. La revisión se apoya en fuentes publicadas hasta 2025 y en la literatura más amplia de educación médica. La evidencia muestra que la inteligencia artificial puede apoyar el aprendizaje y dar retroalimentación, pero los estudios son pequeños, heterogéneos y mayormente de corto plazo, y la validez de los modelos de lenguaje para la evaluación es irregular. Los principales desafíos son curriculares: decidir qué enseñar, cómo enseñarlo y cómo preservar la integridad de la evaluación ante las herramientas generativas. La revisión concluye que la inteligencia artificial debe integrarse en los currículos odontológicos como una herramienta de uso crítico, sin abrazarla acríticamente ni excluirla.

Palabras clave: inteligencia artificial; educación en odontología; currículo; modelos de lenguaje; tecnología educativa; estudiantes de odontología.

INTRODUCTION

Dental education is a long, structured process in which a student moves from basic science to preclinical simulation to patient care, and at every stage the student depends on feedback that is scarce, expensive, and unevenly distributed. Artificial intelligence has begun to enter this process, first as simulation and e-learning, then as image-based tutoring in radiology, and most recently as large language models that can answer questions, generate cases, and draft text. The technology has the potential to address the central scarcity in dental education, which is individualized feedback, but it also introduces new problems for assessment and for the definition of competence.^(1,2) The scarcity of feedback is not incidental but structural: a single instructor cannot observe every student's every attempt, and the consequence is that students practice with less correction than the task requires, which is exactly the gap that a feedback-giving machine is suited to fill. The question that frames this review is therefore not whether the technology should enter dental education, which is already answered, but what it should be made to do there and how its presence should change the teaching and the assessment that surround it.

The question of what artificial intelligence means for dental education is distinct from what it means for dental practice. In practice, the question is whether a model can diagnose or plan correctly; in education, the question is whether a model can help a student learn to diagnose and plan, and whether the presence of the model changes what it means to have learned. These are different questions, and the education literature has addressed them unevenly, with more attention to the novelty of the tools than to the learning they produce.^(3,4) A model that is excellent in practice can still be a poor teacher, if it produces answers without explanations or explanations without the errors from which a student learns, and a model that is imperfect in practice can still be a good teacher, if its errors are the occasion for the student's own reasoning. The literature has not yet internalized this asymmetry, and the result is that the educational value of the tools is judged by the standards of the practice literature, which is the wrong standard.⁽⁴⁾

The rapid arrival of large language models has sharpened the question. A student can now ask a model to answer an examination question, draft a treatment plan, or write a reflective essay, and the educator must decide whether this is cheating, assistance, or something in between. The distinction is not semantic, because it determines whether the curriculum treats the tool as a threat to be policed or a resource to be taught.^(5,6) The arrival was sudden enough that the institutions did not have time to decide before the students had begun to use the tools, and the policies have been, in many places, a response to a practice that had already spread rather than a framework for it.⁽⁵⁾ The question of what counts as cheating is also shifting, because the boundary between the tool as an aid and the tool as a substitute is drawn differently for a grammar check than for a diagnosis, and the same act can be legitimate in one context and illegitimate in another.⁽⁷⁾

The aim of this narrative review is to synthesize the applications, evidence, and curricular challenges of artificial intelligence in dental education, organized chronologically by the stage of adoption, and to identify what the evidence supports and what remains open. The review is written for educators and curriculum planners who must decide how to respond to a tool that their students are already using.⁽⁸⁾ The chronological organization is chosen deliberately, because the field's history is short and its trajectory is the clearest evidence of where it is going, and because the stages map onto a progression from tools that fit the existing curriculum to tools that force the curriculum to change. The review does not attempt to rank or recommend specific tools, which would be obsolete before it was printed, but instead to identify the stable questions that any tool, present or future, raises for teaching and assessment.

METHODS

This review followed the principles of the Scale for the Assessment of Narrative Review Articles, and it did not follow a systematic review protocol.⁽⁹⁾ The sources were bibliographic databases searched, using queries that combined terms for artificial intelligence, large language models, and ChatGPT with dental education, students, curriculum, and assessment, with a publication window of 2018 to 2025 for the dental studies and no lower limit for the conceptual and methodological literature. Selection was by relevance to one of the four stages of adoption, and the synthesis is narrative and organized chronologically. The searches returned a large body of work on artificial intelligence in dental practice, from which the subset addressing education was retained, and the medical education literature was searched separately to supply the conceptual and methodological references that the dental literature does not yet provide.

The four stages were defined before the synthesis: simulation and e-learning, image-based artificial intelligence in radiology teaching, large language models and generative artificial intelligence, and assessment integrity and curricular integration. These stages overlap rather than succeed one another, and a study is assigned to the stage in which its contribution lies, but the chronological frame is useful because it tracks how the field has moved from the tools it already understood to the tools it is still learning to understand. The stages and their challenges are summarized in a single table (Table 1).

Table 1. Stages of artificial intelligence in dental education, with their applications and challenges.

Stage	Applications	Evidence	Main challenge
Simulation and e-learning	Automated feedback on preparations, virtual patients	Small positive studies	Transfer to clinical competence
Image-based AI in radiology	Automated interpretation feedback	Small heterogeneous studies	Whether the model teaches or answers
Large language models	Tutoring, question answering, case generation	Uneven accuracy across tasks	A useful tutor and a misleading authority
Assessment and curriculum	Integrity, competency definition	More proposals than evidence	Measuring competence with and without the model

Note: The stages overlap and are presented in the chronological order in which they entered the field.

The chronological frame should not be read too rigidly, because the stages coexist in the present curriculum, and a contemporary dental school may have all four at once, in different courses and at different levels of maturity.

DEVELOPMENT

Simulation and e-learning

The earliest and least controversial use of artificial intelligence in dental education is in simulation and e-learning, where it extends techniques that predate the current wave of interest. Automated assessment of preclinical preparations, virtual patients, and adaptive e-learning systems have all been explored, and their common feature is that they provide feedback at a scale that a human instructor cannot.(10,11) The continuity with earlier educational technology is worth emphasizing, because it means this stage does not raise the new questions that the later stages do; it simply does the old things better and at greater scale, and it can therefore be adopted without reopening the fundamental questions of the curriculum that the later stages will force, and this is precisely why it is, and should properly remain, the least controversial part of the field, and the part whose adoption the currently available empirical evidence supports most clearly. The value of these systems is established in principle, and the dental literature has begun to report their use in specific skills, such as cavity preparation and crown preparation, where an automated system can measure the dimensions of the student's work and return a score.(12)The measurement is the key: a preparation has objective dimensions, and an automated system can compare the student's work against a standard with a precision and a consistency that a human eye, however experienced, cannot match.

The evidence for these systems is positive but thin. Studies report that automated feedback improves performance on the specific task and that students value the immediate feedback, but the studies are small, single-institution, and short-term, and few of them measure whether the improvement transfers to later clinical performance.(12,13) The gap between task-level improvement and clinical transfer is the recurring limitation of this literature, and it applies equally to the later stages. The reason the gap matters is that the goal of the preclinical stage is not the preparation itself but the skill that the preparation exercises, and a system that improves the preparation without improving the skill has not achieved the educational purpose, however well it scores the task.(13)

The design of these systems also carries an implicit pedagogical choice. A system that scores only the final preparation teaches the student to optimize the outcome, whereas a system that scores the process, the sequence and the technique, teaches the student to optimize the approach, and the two are not the same. The literature has attended more to the measurement of the outcome than to the pedagogy of the process, and this is a direction in which the simulation stage

could mature.(14) The distinction between outcome and process scoring is not academic, because the clinical skill is a process, and a student who has learned to produce a good preparation by a bad process has not learned the skill that will transfer to a patient, even though the preparation itself is acceptable.(13)

The simulation stage is nonetheless the most mature part of the field, because the tools are well understood, the outcomes are measurable, and the risks are low. The remaining question is not whether artificial intelligence can support simulation but whether the dental schools will invest in it, and the answer is shaped by cost and infrastructure as much as by evidence.(15)The applications described in this stage are the ones least disrupted by the later arrival of generative models, because a feedback system that measures a preparation is not replaced by a conversational model, and the two coexist as complementary tools rather than as competitors.(14,16) The distinction between the two is worth keeping clear: the simulation system gives feedback on a motor and spatial skill that it can measure, while the conversational model gives feedback on knowledge and reasoning that it can only approximate, and the confidence of the former is earned in a way that the confidence of the latter is not.

Image-based artificial intelligence in radiology education

The second stage is the use of image-based artificial intelligence in the teaching of dental and maxillofacial radiology, which is the part of the curriculum where the models are closest to their clinical use. Students are taught to interpret radiographs, and the models that assist this interpretation are the same models that the students will encounter in practice, so the educational question is how to use them to teach rather than merely to answer.(17,18) This stage has a natural advantage over the first, because the educational object, the radiograph, and the clinical object are the same, and a model trained for diagnosis can be repurposed for teaching with comparatively little additional work.(17)The stage also has a natural urgency, because radiograph interpretation is a skill in which the consequences of error are immediate and serious, and the pressure to teach it well is correspondingly high.(18)

The models in this stage vary in their suitability for teaching. A model that localizes a finding and returns a confidence is well suited to teaching, because the student can compare the finding and the confidence against their own reading, whereas a model that returns only a classification is less suited, because it offers the answer without the evidence for it.(17)The best educational uses of these models are therefore the ones in which the model's reasoning is visible, and the worst are the ones in which the model is a black box that the student is told to trust, which is the same distinction that governs the clinical use of the same models. The radiology classroom is also where the student first meets the model as a colleague rather than as a textbook, and the habits formed there, of trusting, questioning, or verifying the model, are the habits the student will carry into practice, which makes this stage a formative one beyond the specific skill of radiograph interpretation.(19) The radiology course is therefore the natural place to begin the critical digital literacy that the later stages will demand, and the educators in this stage are, whether they intend it or not, teaching the student what it means to work with a machine that also reads the image.

The applications include automated feedback on radiograph interpretation, systems that flag findings for the student to confirm, and comparisons between the student's reading and the model's reading. The evidence suggests that such systems can support learning and can help students calibrate their own interpretation, but the studies are again small and heterogeneous, and the comparison between the model and the student is complicated by the fact that the model is in several of the studies the reference against which the student is measured.(20,21) The comparison between the student and the model is itself pedagogically useful, because it makes the student's own reading an object of reflection rather than a final answer, and a student who must justify a reading that disagrees with the model is practicing the very reasoning that the curriculum aims to teach.(19,21)

The deeper question in this stage is whether the model should be visible to the student during learning or hidden. A model that flags findings before the student has attempted the interpretation may teach the student to rely on the flag rather than to see the finding, whereas a model that confirms the student's interpretation afterwards may teach the student to see.(3)The pedagogy of this distinction is not yet worked out, and it is a question that the clinical and the educational uses of the same model raise in parallel. The distinction between scaffolding and dependence is old in education, but the speed and confidence of the model give it a new sharpness, because a student can come to rely on the flag so quickly that the underlying skill never forms.(19,22)

Large language models and generative artificial intelligence

The third and most disruptive stage is the arrival of large language models, which differ from the earlier tools in being general-purpose, conversational, and capable of producing text and reasoning on demand. A dental student can now ask a model to explain a concept, answer a multiple-choice question, generate a differential diagnosis, or draft a treatment plan, and the model will produce a plausible answer for all of these.(6,23) The generality is what makes the stage disruptive, because the earlier tools were each built for a single task and stayed within it, whereas the language model crosses every boundary and appears wherever text or reasoning is produced, which in a dental curriculum is nearly everywhere.

The arrival also changed the relationship between the student and the tool. The earlier tools required the student to learn an interface and to operate within a defined task, which itself taught something about the task, whereas the language model requires nothing of the student but a question, and the effort that the interface once demanded is gone. This ease is both the model's attraction and its danger, because the effort that the student does not make is the effort that the learning depends on.(6)The educator's task is therefore to reintroduce the effort, by asking the student to verify, to justify, and to extend the model's output, rather than to accept it.(22)The effort is not reintroduced for its own sake but because it is the mechanism of learning, and a tool that removes the effort removes the learning along with it, unless the educator deliberately puts the effort back.

The evidence on the accuracy of these models in the dental domain is mixed and uneven across tasks. Studies of large language models on dental examination questions report accuracy that varies with the subject, the question format, and the model, with stronger performance on recall and weaker performance on reasoning and on questions that require integration across a clinical scenario.(24-26)The subject-level variation is consistent enough to be informative: the models do well on the basic sciences, where the answer is a single fact, and worse on the clinical subjects, where the answer depends on weighing several findings, which is the same pattern observed across medical education more broadly.(24)The same unevenness appears in the medical literature, where the models perform well on some licensing questions and poorly on others, and where the accuracy is sensitive to prompting and to the specific model.(27,28) The dental-specific evaluations add a further layer, because the dental examination includes both knowledge questions and case-based questions, and the models handle the former better than the latter.(29,30)

The educational significance of this unevenness is that a model can be a useful tutor and a misleading authority at the same time. A student who asks the model to explain a concept may receive an accurate explanation, while a student who asks the model to confirm a diagnosis may receive a confident but wrong answer, and the model does not distinguish these cases for the student.(11,19) The task of the educator is therefore to teach the student to use the model as a tool whose output must be verified, which is a new form of digital literacy that the curriculum does not yet contain.(22)The confidence of the model is the central hazard, because a wrong answer delivered with confidence is more dangerous to a learner than a wrong answer delivered with uncertainty, and the model's default is confidence regardless of correctness.(19)The student must therefore be taught a habit that runs against the model's design, namely to treat every output as a claim to be checked rather than a fact to be accepted, and this habit is the core of the digital literacy that the curriculum needs to build.

The potential of the model as a tutor is nonetheless real. The model is always available, never impatient, and able to explain the same concept in as many ways as the student asks for, which are properties that no human tutor possesses in combination.(23)The pedagogical literature distinguishes the model's use as a source of explanations, where its unevenness is tolerable because the student can compare and verify, from its use as a source of answers, where its unevenness is not tolerable because the student cannot tell the difference.(23,26) The curriculum should therefore teach the former use and caution against the latter. The distinction is teachable, and a student who is trained to ask the model for explanations rather than answers, and to verify each explanation against the course material, has acquired a durable skill that will outlast the specific model.(22)

Assessment integrity and curricular integration

The fourth stage is the most difficult, because it concerns what the presence of these tools means for assessment and for the curriculum as a whole. The immediate problem is integrity: a student who submits a model-generated answer or essay has not demonstrated the competence that the assessment was meant to measure, and the distinction between legitimate assistance and illegitimate delegation is hard to draw and harder to enforce.(5,7)

The difficulty is that the same tool can be used in both ways, as an aid to thinking and as a substitute for it, and the difference lies in the student's intent, which an assessment cannot see. (7) The technical tools that claim to detect model-generated text are unreliable, and the reliance on them is a losing strategy, because they produce both false accusations and false reassurance and because the models that evade them improve faster than the detectors do. (7) The sustainable response is therefore not detection but design, in which the assessment is structured so that the model's contribution, whether present or not, does not undermine what the assessment measures.

The deeper problem is curricular. If a model can answer the examination questions, then the examination questions are testing what the model can do rather than what the clinician must be able to do, and the assessment must change accordingly. The responses proposed in the literature range from returning to proctored, in-person, and oral assessment, to redesigning the questions so that they require what the model cannot do, to accepting the model as part of the assessment and grading the student's use of it. (5,31,32) The integrity problem is made concrete by studies that assess how well the models perform on the actual dental assessments, which shows that the concern is not hypothetical but already realized. (33,34) Each of the proposed responses has a cost: proctored and oral assessment is expensive and does not scale, redesigned questions are difficult to write and to maintain, and accepting the model in the assessment requires a standard for what constitutes good use of it, which is itself a new competency to define and to teach.

The curricular question is what to teach about artificial intelligence itself. Students will practice alongside these tools, and they need to understand what the tools do, where they fail, and how to use them responsibly, and this understanding is a new competency that the curricula are only beginning to define. (1,2) The surveys of students and educators show high awareness and cautious attitudes, with students generally more enthusiastic and educators more guarded, and the gap between the two is itself a curriculum problem. (2,35,36) The surveys also reveal that knowledge about the tools is unevenly distributed, with many students reporting exposure to the tools without a corresponding understanding of their limitations, which suggests that exposure alone is not competence. (37,38) The proposals for a defined artificial intelligence literacy in dental education, and for its assessment, are the concrete form of this new competency. (39,40)

The definition of this competency is itself a matter of debate. At a minimum it includes knowing what the tools are, where they fail, and how to verify their output, and beyond that minimum the proposals diverge, with some adding a technical understanding of how the models work and others adding an ethical understanding of when they should be used. (39) The divergence is healthy, because the competency is new and the field has not settled on its boundaries, but it also means that two schools that both claim to teach artificial intelligence literacy may be teaching quite different things, and a shared definition would make the curricula comparable. (1,40) The shared definition, when it comes, will need to be specific enough to be taught and assessed but general enough to survive the rapid change of the underlying tools, and finding that balance is a task the field has only begun.

DISCUSSION

Across the four stages, a single trajectory is visible. The field has moved from tools that extend existing educational methods, in simulation, to tools that challenge the methods themselves, in generative models, and the pace of that movement has outpaced the evidence. The simulation and image-based stages have a body of small positive studies, while the language model stage has a body of accuracy studies that is growing faster than the studies of learning, and the assessment stage has more proposals than evidence. (41,42) This trajectory mirrors the broader medical education literature, where the same lag between tool availability and learning evidence has been observed. (22,27) The lag is not a failure of the field but a property of the sequence in which any new tool arrives: first it is demonstrated, then it is adopted, and only then is it studied, and the dental field is still in the middle stage, where adoption outruns evidence. The lag has a consequence for practice, which is that the curricular decisions are being made now, under uncertainty, and the educators cannot wait for the evidence that would make the decisions easy, because the students are using the tools now and the curriculum must respond now.

The central gap is the absence of evidence that connects the tool to the learning. The studies measure whether a model is accurate or whether a student's performance improves on a task, but they rarely measure whether the student learns something that persists, transfers, and changes their later clinical behavior. Until that link is established, the field cannot distinguish a tool that helps students learn from a tool that merely helps them pass. (43,44) The distinction matters because a

tool that helps a student pass without learning is, in an educational sense, worse than no tool at all, since it defers the failure to the point where the patient is involved.(7,28)

The evidence that does exist is suggestive but not decisive. The studies that measure a learning outcome tend to report a benefit, and the studies that measure only the model's accuracy tend to report the model's accuracy, and the two bodies of work do not yet meet in the middle, where a model's accuracy is linked to a student's learning.(43) Bridging them requires studies that are harder to design and slower to run, because learning is a slower and noisier outcome than accuracy, and the field has preferred the faster and cleaner measurement, which is understandable but leaves the central question unanswered.(41,44) The preference is also a consequence of the incentives: an accuracy study is quick to publish and easy to review, while a learning study is slow and methodologically demanding, and the field's incentives have therefore favored the measurement that is least informative about education.(43)

The assessment question is the most urgent, because it is the one that the arrival of the tools has already forced. The profession must decide, and is deciding, whether the model is a threat to the integrity of assessment or an opportunity to make assessment more authentic, and the answer determines the shape of the whole curriculum. The most coherent position in the literature is that assessment should measure what a clinician must be able to do without the model, and separately measure what the clinician can do with the model, because both are part of future practice, and neither should be confused with the other.(5,31) This two-part assessment is the only position that is honest about the future, in which the clinician will always have the model available, and about the present, in which the student must still demonstrate the competence that the model cannot supply.(31) The two-part assessment is not a compromise but a more complete account of competence, because the clinician of the future will be judged on both the skill and the skill at using the tools, and the assessment should mirror that reality rather than resist it.

The integration of the tools into the curriculum is not a single decision but a sequence of them, each of which the institution must make for itself. Whether to teach the tools, in which courses, to what depth, and with what stance toward their use, are all open questions that the literature can inform but not answer, because the answers depend on the institution's mission, its students, and its resources. The literature provides the options and the considerations, and the institutions provide the decisions, and the result is that the curricula are diverging rather than converging, which is both a risk and an experiment from which the field can learn.(8,32) The divergence is most visible in the stance toward the tools, which ranges from permissive to restrictive, and the stance, more than any specific course content, is what shapes the students' eventual relationship with the technology.(8)

The limitations of this review are those of a young and heterogeneous literature. The dental studies are mostly small, single-institution, and short-term, and the review therefore leans on the medical education literature for the conceptual frame. The selection was narrative rather than systematic, and the rapid pace of publication means that the account is necessarily a snapshot. The review also cannot assess the long-term outcomes of the curricula that are only now being designed, because those outcomes do not yet exist. These limitations are themselves informative: they show that the field is at the stage of early adoption, where the questions are being posed faster than the evidence can answer them, and where any synthesis is provisional by necessity rather than by choice. The review should therefore be read as an organizing account of a moving field rather than as a settled summary, and its conclusions are offered with the caution that a rapidly changing technology deserves.

The agenda that follows is concrete. For research, the field needs studies that measure learning and transfer, not merely accuracy, and that compare the tool against the alternative rather than against nothing. For the curriculum, the field needs a defined competency for artificial intelligence literacy, a pedagogy for using the tools critically, and an assessment framework that separates competence without the tool from competence with it. For the profession, the field needs a shared position on academic integrity that is principled rather than reactive.(8,45)

The agenda also has a temporal dimension. The tools will change, and a curriculum that teaches a specific model or a specific interface will be obsolete quickly, whereas a curriculum that teaches the underlying skill, of using an artificial intelligence critically and verifying its output, will remain relevant as the tools evolve.(8) The competency to be taught is therefore not a knowledge of the current tools but a habit of mind, and the institutions that teach the habit rather than the tool will be the ones whose graduates are prepared for whatever the tools become.

The review closes with a note on responsibility. The decision to integrate or to exclude these tools is not value-neutral, because the students will use them either way, and the choice is therefore between teaching their use and leaving it untaught.(8) A curriculum that excludes the tools does

not prevent their use; it merely ensures that the use happens without guidance, and a curriculum that integrates them takes responsibility for a practice that is already occurring. The profession should therefore choose integration, not because the tools are good, but because the alternative is to abandon the students to them, and that abandonment is the one response that the evidence does not support. The integration should be critical rather than celebratory, teaching the student to use the tool and to doubt it at the same time, and the balance between those two teachings is the substance of the curriculum that the field must now build.

CONCLUSIONS

Artificial intelligence has entered dental education in stages, from simulation to image-based tutoring to large language models, and each stage has moved the field closer to a tool that challenges the curriculum itself rather than merely serving it. The evidence supports the use of these tools to provide feedback and to support learning, but it is small and short-term, and it does not yet establish transfer to clinical competence. The accuracy of large language models is uneven across dental tasks, which makes them useful tutors and unreliable authorities at once. The central challenges are curricular and assessment-related: deciding what to teach about the tools, how to preserve assessment integrity, and how to measure competence in a world where the model can answer the questions. The field should integrate artificial intelligence as a tool to be used critically, teaching students both what it can do and how to verify what it produces, rather than either embracing it uncritically or excluding it. The students will use these tools regardless, and the choice the profession actually faces is not whether they are used but whether their use is guided, which is a choice that the curriculum, and not the technology, must make. The curriculum of the future will be judged, in part, by how well it prepares its graduates to work alongside machines that are themselves still learning, and that preparation is the task to which this review points.

REFERENCES

1. Ahmad Satmi AS, Reza N, Khamis MF, Mohamad N, Abdullah JY, Mohd Ali Hanafiah MI, et al. Educational aspects of artificial intelligence in oral and maxillofacial radiology: insights from a scoping review. *BMC Med Educ.* 2025;26(1):164. doi: 10.1186/s12909-025-08493-w
2. Ghaffar RSA, Malik R, Amjad M, Nadeem R, Admani MS, Sanaullah A, et al. Awareness, usage, and perspectives on ChatGPT in dental education among graduate and undergraduate students. *Sci Rep.* 2025;16(1):3747. doi: 10.1038/s41598-025-33704-1
3. Huang S, Qu Y, Sun Y, Chen S, Gu L. Application and development of artificial intelligence and immersive technologies in dental education: a scoping review. *BMC Med Educ.* 2025;26(1):89. doi: 10.1186/s12909-025-08420-z
4. Brondani M, Alves C, Ribeiro C, Braga MM, Garcia RCM, Ardenghi T, et al. Artificial intelligence, ChatGPT, and dental education: Implications for reflective assignments and qualitative research. *Journal of Dental Education.* 2024;88(12):1671-1680. doi: 10.1002/jdd.13663
5. Sallam M, Salim N, Barakat M, Al-Tammemi A. ChatGPT applications in medical, dental, pharmacy, and public health education: A descriptive study highlighting the advantages and limitations. *Narra J.* 2023;3(1):e103. doi: 10.52225/narra.v3i1.103
6. Ali K, Barhom N, Tamimi F, Duggal M. ChatGPT—A double-edged sword for healthcare education? Implications for assessments of dental students. *Eur J Dental Education.* 2024;28(1):206-211. doi: 10.1111/eje.12937
7. Ahmed WM, Azhari AA, Alfaraj A, Alhamadani A, Zhang M, Lu CT. The Quality of AI-Generated Dental Caries Multiple Choice Questions: A Comparative Analysis of ChatGPT and Google Bard Language Models. *Heliyon.* 2024;10(7):e28198. doi: 10.1016/j.heliyon.2024.e28198
8. Verghese BG, Iyer C, Borse T, Cooper S, White J, Sheehy R. Modern artificial intelligence and large language models in graduate medical education: a scoping review of attitudes, applications & practice. *BMC Med Educ.* 2025;25(1):730. doi: 10.1186/s12909-025-07321-5
9. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ.* 2021;n71. doi: 10.1136/bmj.n71
10. Paranjape K, Schinkel M, Nannan Panday R, Car J, Nanayakkara P. Introducing Artificial Intelligence Training in Medical Education. *JMIR Med Educ.* 2019;5(2):e16048. doi: 10.2196/16048
11. Yan L, Sha L, Zhao L, Li Y, Martinez-Maldonado R, Chen G, et al. Practical and ethical challenges of large language models in education: A systematic scoping review. *Brit J Educational Tech.* 2024;55(1):90-112. doi: 10.1111/bjet.13370
12. Thurzo A, Strunga M, Urban R, Surovková J, Afrashtehfar KI. Impact of Artificial Intelligence on Dental Education: A Review and Guide for Curriculum Update. *Education Sciences.* 2023;13(2):150. doi: 10.3390/educs-c13020150
13. Uribe SE, Maldupa I, Kavadella A, El Tantawi M, Chaurasia A, Fontana M, et al. Artificial intelligence chatbots and large language models in dental education: Worldwide survey of educators. *Eur J Dental Education.* 2024;28(4):eje.13009. doi: 10.1111/eje.13009
14. Claman D, Sezgin E. Artificial Intelligence in Dental Education: Opportunities and Challenges of Large Language Models and Multimodal Foundation Models. *JMIR Med Educ.* 2024;10:e52346-e52346. doi: 10.2196/52346
15. Luo J, Peng Q, Cao H, Li W, Tan S. Medical undergraduate students' readiness and anxiety toward artificial intelligence: a systematic review and meta-analysis. *BMC Med Educ.* 2025;26(1):44. doi: 10.1186/s12909-025-08388-w
16. Motaghi A, Sarkhosh S, Nazer A, Hasanzadeh I. Advancement of Artificial Intelligence in Dental Education: A Narrative Review. *Open Access Res J Biol Pharm.* 2025;15(2):006-011. doi: 10.53022/oarjbp.2025.15.2.0045
17. Kiyani A, Hanif F, Muhammad M, Iqbal S, Zaib N, Bashir U, et al. Benchmarking ChatGPT-generated multiple-choice questions against faculty-authored items in dental education. *Sci Rep.* 2025;15(1):44805. doi: 10.1038/s41598-025-28492-7
18. Batgerel OE, Akkitap MP, Sasany R. Diagnostic competence of senior dental students in detecting caries on panoramic radiographs with and without artificial intelligence assistance: a cross-sectional study. *BMC Med Educ.* 2025;25(1):1662. doi: 10.1186/s12909-025-08269-2
19. Huang S, Wen C, Bai X, Li S, Wang S, Wang X, et al. Exploring the Application Capability of ChatGPT as an Instructor in Skills Education for Dental Medical Students: Randomized Controlled Trial. *J Med Internet Res.* 2025;27:e68538. doi: 10.2196/68538
20. Alpay S, Darafarin Y, Dagdelen B, Al-Shammari SMM, Buyukbayram IK. Comparison of ChatGPT-4o and expert evaluation in endodontic education: a cross-sectional pilot study. *BMC Med Educ.* 2025;26(1):12. doi: 10.1186/s12909-025-08358-2
21. Koyuncuoglu CZ, Selcuker AH, Ozyilmaz E. The reliability of answers from four different AI chatbots on periodontology theoretical exam questions: an evaluation in dental education. *BMC Oral Health.* 2025;26(1):114. doi: 10.1186/s12903-025-07387-z
22. Tran M, Balasooriya C, Jonnagaddala J, Leung GKK, Mahboobani N, Ramani S, et al. Situating governance and regulatory concerns for generative artificial intelligence and large language models in medical education. *npj Digit Med.* 2025;8(1):315. doi: 10.1038/s41746-025-01721-z
23. Huang H, Zheng O, Wang D, Yin J, Wang Z, Ding S, et al. ChatGPT for shaping the future of dentistry: the potential of multi-modal large language model. *Int J Oral Sci.* 2023;15(1):29. doi: 10.1038/s41368-023-00239-y
24. Künzle P, Paris S. Performance of large language artificial intelligence models on solving restorative dentistry and endodontics student assessments. *Clin Oral Invest.* 2024;28(11):575. doi: 10.1007/s00784-024-05968-w

25. Shete A, Shete M, Chavan M, Channe P, Sapkal R, Buva K. Evaluation of ChatGPT as a New Assessment Tool in Dental Education. *Journal of Indian Academy of Oral Medicine and Radiology*. 2024;36(3):259-263. doi: 10.4103/jiaomr.jiaomr_62_24
26. Jeon EY, Ahn HN, Lee JH. Evaluating the Feasibility of ChatGPT in Dental Morphology Education: A Pilot Study on AI-Assisted Learning in Dental Morphology. *J Dent Hyg Sci*. 2024;24(4):309-315. doi: 10.17135/jdhs.2024.24.4.309
27. Abd-alrazaq A, AlSaad R, Alhuwail D, Ahmed A, Healy PM, Latifi S, et al. Large Language Models in Medical Education: Opportunities, Challenges, and Future Directions. *JMIR Med Educ*. 2023;9:e48291. doi: 10.2196/48291
28. Nouri H, Mahdavi A, Abedi A, Mohammadnia A, Hamedan M, Amanzadeh M. Performance of large language models in medical licensing examinations: a systematic review and meta-analysis. *J Educ Eval Health Prof*. 2025;22:36. doi: 10.3352/jeehp.2025.22.36
29. Liu M, Okuhara T, Huang W, Ogihara A, Nagao HS, Okada H, et al. Large Language Models in Dental Licensing Examinations: Systematic Review and Meta-Analysis. *International Dental Journal*. 2025;75(1):213-222. doi: 10.1016/j.identj.2024.10.014
30. Başkan HK, Başkan B. Performance comparison of large language models on pediatric dentistry questions in the Turkish dentistry specialization examination. *BMC Med Educ*. 2025;25(1):1734. doi: 10.1186/s12909-025-08315-z
31. Artsi Y, Sorin V, Konen E, Glicksberg BS, Nadkarni G, Klang E. Large language models for generating medical examinations: systematic review. *BMC Med Educ*. 2024;24(1):354. doi: 10.1186/s12909-024-05239-y
32. Gordon M, Daniel M, Ajiboye A, Uraiby H, Xu NY, Bartlett R, et al. A scoping review of artificial intelligence in medical education: BEME Guide No. 84. *Medical Teacher*. 2024;46(4):446-470. doi: 10.1080/0142159x.2024.2314198
33. Alnsour MM, Al-Ghannam W, Alraheam IA, Sabrah AHA, Oweis Y, AL-Omiri MK. Assessing the suitability of ChatGPT in responding to public inquiries about dental crown restorations. *BMC Oral Health*. 2025;25(1):1949. doi: 10.1186/s12903-025-07281-8
34. Alqahtani H. Assessment of artificial intelligence chatbots in responding to dental occlusion questions: a comparative study. *BMC Oral Health*. 2025;26(1):201. doi: 10.1186/s12903-025-07573-z
35. Yilmaz BE, Gokkurt Yilmaz BN, Ozbey F. Artificial intelligence performance in answering multiple-choice oral pathology questions: a comparative analysis. *BMC Oral Health*. 2025;25(1):573. doi: 10.1186/s12903-025-05926-2
36. Ardila CM, Pineda-Vélez E, Vivares-Builes AM. Artificial Intelligence in Endodontic Education: A Systematic Review with Frequentist and Bayesian Meta-Analysis of Student-Based Evidence. *Dentistry Journal*. 2025;13(11):489. doi: 10.3390/dj13110489
37. Kurt Ö, Şimsek E. Knowledge-level comparison in pulpal and periapical diseases: dental students versus artificial intelligence models (Gemini, Microsoft Copilot, ChatGPT-3.5, ChatGPT-4o): cross-sectional study. *BMC Med Educ*. 2025;25(1):1657. doi: 10.1186/s12909-025-08263-8
38. Ellakany P, Tauqir S, Ali S, Mikhail SS. Artificial intelligence usage among dental students/dentists of different educational level: a multi-country survey. *BMC Oral Health*. 2025;26(1):81. doi: 10.1186/s12903-025-07476-z
39. El-Hakim M, Anthonappa R, Fawzy A. Artificial Intelligence in Dental Education: A Scoping Review of Applications, Challenges, and Gaps. *Dentistry Journal*. 2025;13(9):384. doi: 10.3390/dj13090384
40. Monalisa S, Alipuor M, Paul D, Rahman MA, Siddika N, Apu EH, et al. Transforming Dental Care, Practice and Education with Additive Manufacturing and 3D Printing: Innovations in Materials, Technologies, and Future Pathways. *Dentistry Journal*. 2025;13(12):555. doi: 10.3390/dj13120555
41. Lucas HC, Upperman JS, Robinson JR. A systematic review of large language models and their implications in medical education. *Medical Education*. 2024;58(11):1276-1285. doi: 10.1111/medu.15402
42. Simoni J, Urtubia-Fernandez J, Mengual E, Simoni DA, Royo M, Egaña-Yin D, et al. Artificial intelligence in undergraduate medical education: an updated scoping review. *BMC Med Educ*. 2025;25(1):1609. doi: 10.1186/s12909-025-08188-2
43. Pupic N, Chaffari-zadeh A, Hu R, Singla R, Darras K, Karwowska A, et al. An evidence-based approach to artificial intelligence education for medical students: A systematic review. *PLOS Digit Health*. 2023;2(11):e0000255. doi: 10.1371/journal.pdig.0000255
44. Lin Y, Luo Z, Ye Z, Zhong N, Zhao L, Zhang L, et al. Applications, Challenges, and Prospects of Generative Artificial Intelligence Empowering Medical Education: Scoping Review. *JMIR Med Educ*. 2025;11:e71125-e71125. doi: 10.2196/71125
45. Xu X, Chen Y, Miao J. Opportunities, challenges, and future directions of large language models, including ChatGPT in medical education: a systematic scoping review. *J Educ Eval Health Prof*. 2024;21:6. doi: 10.3352/jeehp.2024.21.6

FUNDING

No financing

CONFLICT OF INTEREST

None

AUTHORSHIP CONTRIBUTIONS

Drafting – original draft: Maylon Felix Guerra Reyes, Ernesto Learsys Campello Trujillo.

Writing–review and editing: Maylon Felix Guerra Reyes, Ernesto Learsys Campello Trujillo.