



Ethics, bias and regulation of artificial intelligence in dentistry

Ética, sesgo y regulación de la inteligencia artificial en odontología

Youris Samón Mayan¹

Marta María Capote Fernández¹ 

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

Submitted: 06-05-2025 | Revised: 02-07-2025 | Accepted: 11-10-2025 | Published: 02-11-2025

Corresponding Author: Marta María Capote Fernández 

ABSTRACT

Artificial intelligence is transforming dentistry, but its adoption raises ethical, bias-related, and regulatory questions that the accuracy-focused literature has left largely unexamined. This narrative review synthesizes the ethical dimensions of dental artificial intelligence across five axes: algorithmic bias and fairness, ethical principles, regulation and governance, data privacy and consent, and accountability and liability. The review draws on sources published up to 2025 and on the broader medical artificial intelligence literature. Dental models inherit demographic and acquisition biases from their training data, and fairness metrics are rarely reported. The principles of autonomy, beneficence, non-maleficence, and justice apply directly, and trustworthy use requires explainability and transparency. Regulatory frameworks are emerging but heterogeneous, and dental tools sit uneasily between medical device and general software regulation. Data governance, informed consent, and the allocation of responsibility for model errors remain unresolved. The review concludes that the clinical value of dental artificial intelligence will depend as much on resolving these dimensions as on further gains in accuracy.

Keywords: artificial intelligence; dentistry; ethics; bias; government regulation; machine learning.

RESUMEN

La inteligencia artificial está transformando la odontología, pero su adopción plantea interrogantes éticos, de sesgo y regulatorios que la literatura centrada en la precisión ha dejado en gran medida sin examinar. Esta revisión narrativa sintetiza las dimensiones éticas de la inteligencia artificial odontológica en cinco ejes: sesgo algorítmico y equidad, principios éticos, regulación y gobernanza, privacidad de datos y consentimiento, y responsabilidad. La revisión se apoya en fuentes publicadas hasta 2025 y en la literatura más amplia de inteligencia artificial médica. Los modelos odontológicos heredan sesgos demográficos y de adquisición de sus datos de entrenamiento, y las métricas de equidad rara vez se reportan. Los principios de autonomía, beneficencia, no maleficencia y justicia se aplican directamente, y un uso confiable requiere explicabilidad y transparencia. Los marcos regulatorios son emergentes y heterogéneos. La gobernanza de los datos, el consentimiento informado y la asignación de responsabilidad por los errores del modelo siguen sin resolverse. La revisión concluye que el valor clínico de la inteligencia artificial odontológica dependerá tanto de resolver estas dimensiones como de nuevas ganancias de precisión.

Palabras clave: inteligencia artificial; odontología; ética; sesgo; regulación gubernamental; aprendizaje automático.

INTRODUCTION

Artificial intelligence has moved rapidly from a research curiosity to a clinical tool across dentistry, with models now proposed or deployed for caries detection, periodontal assessment, cephalometric analysis, tooth segmentation, and treatment planning.^(1,2) The literature on these applications is dominated by accuracy metrics, and the question that drives most studies is whether a model performs as well as or better than a clinician. The ethical, bias-related, and regulatory dimensions of the same technologies have received far less attention, even though they determine whether the tools can be used safely and equitably in practice.⁽³⁾ This imbalance is itself a finding worth noting: the field has invested heavily in making the models accurate and comparatively little in making them fair, lawful, and accountable. The imbalance is not surprising, because accuracy is easy to measure and fairness is not, but it is consequential, because a model is only as useful as it is trustworthy.

The three dimensions are connected rather than independent. Algorithmic bias determines whether a model performs equitably across the populations it serves, ethics determines whether its use respects patient autonomy and promotes benefit, and regulation determines whether its deployment is lawful and accountable. A model that is accurate on average but biased against a subgroup is a clinical hazard, and a model that is accurate and fair but deployed without a regulatory basis is a legal and institutional hazard.^(4,5) These dimensions therefore cannot be separated from the accuracy literature, and a complete assessment of dental artificial intelligence must address all of them. The present review is organized so that each dimension is treated in its own right and then brought together, because the whole point is that they are not separable in practice.

The connection runs in both directions. A biased model is not merely an ethical problem; it is also an accuracy problem, because the bias means the accuracy is unequally distributed. And an unregulated model is not merely a legal problem; it is also a trust problem, because a clinician cannot responsibly rely on a tool whose provenance and validation are unknown. The three dimensions are therefore best understood as three lenses on a single question: whether a dental artificial intelligence tool can be trusted to do what it claims, for every patient, within a lawful and accountable framework.

The aim of this narrative review is to synthesize the ethics, bias, and regulatory dimensions of artificial intelligence in dentistry across five thematic axes, and to identify the tensions and gaps that remain. The review is organized as a synthesis rather than a systematic enumeration, and it draws on the dental literature where it exists and on the broader medical artificial intelligence literature where the dental evidence is thin.⁽⁶⁾ The intended audience is the clinician and the researcher who must decide whether and how to adopt these tools, and the review is written so that the technical, the ethical, and the regulatory considerations can be weighed together rather than in isolation. The five axes are not intended to be exhaustive of the subject, but they capture the questions that arise most consistently when the technology moves from the laboratory into the clinic.

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.⁽⁷⁾ The sources were bibliographic databases searched using queries that combined terms for artificial intelligence with dentistry and with ethics, bias, fairness, and regulation, with a publication window of 2018 to 2025 for the dental studies and no lower limit for the conceptual and methodological literature. Institutional and regulatory documents were included where they established the framework against which the studies were read. Selection was by relevance to one of the five thematic axes, and the synthesis is narrative and organized by axis rather than by study. The searches returned a substantial body of work on artificial intelligence in dentistry in general, from which the subset addressing ethics, bias, and regulation was retained, and the broader medical artificial intelligence literature was searched separately to supply the conceptual and methodological references that the dental literature does not yet provide.

The five axes were defined before the synthesis: algorithmic bias and fairness, ethical principles, regulation and governance, data privacy and consent, and accountability and liability. For each axis, the review synthesizes the state of the evidence, the points of agreement and disagreement, and the remaining gaps. The review does not aim to be exhaustive, and it does not quantify the frequency of any finding; it aims instead to organize a dispersed literature into a coherent account. The five axes and their central claims are summarized in a single table (Table 1).

Table 1. Synthesis of the thematic axes of ethics, bias and regulation of artificial intelligence in dentistry.

Axis	Central claim	Main unresolved question
Algorithmic bias and fairness	Models inherit demographic and acquisition bias and are rarely audited	How to make subgroup reporting standard
Ethical principles	Autonomy, beneficence, non-maleficence and justice apply directly	How to operationalize explainability and oversight
Regulation and governance	Regulatory status is unsettled and heterogeneous	Which dental tools are devices and how to monitor updates
Data, privacy and consent	Training data governance and consent predate machine learning	A consent model specific to machine learning
Accountability and liability	Responsibility for errors is distributed and unclear	Who is accountable when a model errs

Note: The table summarizes the central claim and the main unresolved question of each axis.

DEVELOPMENT

Algorithmic bias and fairness

The central risk in this axis is that a dental model trained on a non-representative dataset will perform unevenly across the populations it serves, and that the unevenness will reproduce or amplify existing inequities.⁽⁴⁾ The medical literature provides the clearest warning: a widely used algorithm in health management systematically under-rated the disease burden of Black patients because it used health spending as a proxy for need, and the bias arose not from the model but from a biased choice of outcome.⁽⁴⁾ The same mechanism applies in dentistry, where models are trained on radiographs and images drawn predominantly from the institutions and populations that produce the most research, and where demographic and acquisition characteristics are rarely balanced.^(8,9) The lesson of the landmark case is that the bias is usually not in the model's code but in the data and in the choices that define the task, which means that auditing the model in isolation, without auditing the data and the outcome, will miss the problem.

The dental literature has begun to document this. Studies of facial and intraoral image analysis have shown that performance varies with skin tone and with demographic characteristics, and that models trained on one population can degrade when applied to another.^(8,10) Fairness has been proposed as an explicit evaluation criterion for dental models, with calls for the reporting of subgroup performance alongside the overall accuracy, but the criterion is rarely applied in practice.⁽¹¹⁾ The fairness literature in medical imaging provides the tools, including subgroup analysis, calibration by demographic, and algorithmic audits, and their translation to dentistry is overdue.^(12,13) Recent work has begun to tackle the problem directly, proposing concrete steps to overcome bias in dental image analysis and arguing that the obstacles are institutional as much as technical.^(14,15)

The consequences of ignoring this axis are not hypothetical. A model that is less accurate for a particular skin tone or a particular dental morphology will misdiagnose those patients at a higher rate, and because the affected groups are precisely those that already face barriers to care, the model converts a data artifact into a health disparity. The mitigation is to report subgroup performance, to audit for bias before deployment, and to correct the training data where it is unbalanced, but these steps require the dental community to treat fairness as a first-order requirement rather than an afterthought.^(11,16)

The bias in dental models has several distinct sources, and distinguishing them matters for remediation. The first is representation bias, in which the training data under-represent a population and the model therefore performs poorly on it; this is the most commonly discussed and the most tractable, because it can be addressed by collecting more diverse data.⁽⁹⁾ The second is label bias, in which the reference standard itself is uneven, as when the expert annotation is less reliable for one group than another; this is harder to detect because it is embedded in the ground truth.⁽¹⁰⁾ The third is measurement bias, in which the image acquisition differs across settings and the model, trained on one setting, misreads another. All three can coexist, and an audit that finds only the first may miss the others, which is why a fairness evaluation needs to probe each source rather than rely on a single metric. The three sources also require different remedies: representation bias requires more diverse data, label bias requires a better reference, and measurement bias requires

calibration across acquisition settings, and a remedy aimed at one will not fix the others.

Ethical principles

The four classical principles of autonomy, beneficence, non-maleficence, and justice apply directly to dental artificial intelligence, and each raises questions that the accuracy literature does not answer.(17,18) Autonomy is engaged because a patient has a right to know that a machine contributed to their diagnosis and to decline it, which requires a process of consent that many dental practices do not yet have. Beneficence and non-maleficence are engaged because the tool must be shown to improve outcomes without introducing new harms, and the harms of a misdiagnosis are borne by the patient, not by the model.(19)The principle of non-maleficence is the most immediate of the four, because the harm of an incorrect model output is direct and clinical, and the burden of showing that the harm is outweighed by the benefit rests on those who deploy the tool, not on those who are subjected to it.

These principles are not new, but their application to artificial intelligence forces a re-examination of assumptions that the profession has long taken for granted. The assumption that the clinician alone is responsible for a diagnosis, for example, becomes harder to maintain when a model has contributed to it, and the assumption that a patient's consent covers the use of their data becomes harder to maintain when the data are reused for training. The value of the classical principles is precisely that they are old enough and general enough to guide these new questions, and the task is to apply them with the specificity that the technology requires rather than to invent new principles.(17)

The principle of justice is the most structurally demanding, because it asks whether the benefits of the technology are distributed equitably and whether its errors fall on the already disadvantaged. The ethical frameworks proposed for dental artificial intelligence emphasize the need for transparency, explainability, and human oversight, and they converge on the view that the model should be a decision support rather than a decision maker.(20,21) The role of the clinician in this arrangement is central, because the ethical responsibility ultimately rests with the human who interprets the model, and the value of the tool depends on that human being equipped to exercise it.(22,23)

The ethical principles also have a temporal dimension. A model that is ethical at deployment may become less so as the population changes, as the acquisition equipment changes, or as the model is updated, and the ethics of the tool cannot be certified once and then forgotten.(20)The same is true of the clinician's obligation: the duty to use the tool well is not discharged by a single training session but is an ongoing responsibility that must keep pace with the tool's evolution. This temporal dimension, the fact that both the tool and the context are moving, is one reason the static frameworks inherited from medical device regulation fit artificial intelligence so poorly, and it is a recurring theme across all five axes. Explainability is not merely a technical convenience; it is the condition under which a clinician can responsibly accept or override a model's recommendation, and its absence converts the clinician into a passive conduit for an opaque system.(21)

The ethical dimension also extends to the scholarly process itself. The use of artificial intelligence in the authorship of scientific articles in dentistry has raised questions of attribution, integrity, and accountability, and the journals have begun to formulate policies that the individual author must navigate.(24)The same concern applies to the clinical record, where a model-generated interpretation must be traceable to a responsible clinician, and where the provenance of the text or the measurement must be transparent to anyone who reads it later.(24,25)

The ethics of the scholarly use of artificial intelligence are a microcosm of the ethics of the clinical use. In both, the question is the same: what is the human responsible for, and what can be delegated to the machine without transferring the responsibility along with the work. An author who uses a model to draft a section of a manuscript remains responsible for the accuracy of the section, just as a clinician who uses a model to interpret a radiograph remains responsible for the diagnosis, and the failure to recognize this transfer of responsibility, rather than the use of the tool itself, is the ethical failure.(24,26) The principle generalizes: artificial intelligence changes the division of labor, but it does not change the locus of responsibility, and the ethical frameworks that govern the profession must be read in that light.

Regulation and governance

The regulatory status of dental artificial intelligence is unsettled, and the tools sit uneasily between the categories that regulators have established. Some dental models are marketed as medical devices and fall under the device frameworks of the Food and Drug Administration or

the European Medical Device Regulation, while others are marketed as general software and fall outside them, and the boundary between the two is not always clear.(5,27) The emerging artificial intelligence regulations, such as the European approach of risk-based classification, add a further layer that is only beginning to be applied to dentistry.(28,29) A diagnostic model that flags caries, a planning model that segments teeth, and a conversational model that answers patient questions each occupy a different position on this spectrum, and the same word, artificial intelligence, covers tools with very different regulatory implications.

The regulatory question is complicated by the pace of the technology. The frameworks assume that a device is fixed and can be validated once, whereas a learning model that updates after deployment, or a generative model that produces new outputs for each input, does not fit that assumption.(3)The regulatory gap is therefore not merely administrative; it is structural, and it means that a clinician may deploy a tool whose regulatory status is uncertain and whose post-market behavior is not monitored.(30)The gap is most acute for the generative models, whose outputs are not fixed even for a fixed input, and whose behavior cannot be characterized by a single validation study in the way that a diagnostic threshold can, which places them at the very edge of what the current frameworks were designed to govern.(3)The question of where responsibility should fall, and of how the regulatory burden should be shared between developers and clinicians, is the subject of active debate, with proposals ranging from stricter device classification to professional self-regulation.(31-35)

The governance of dental artificial intelligence is also distributed across institutions. The clinic, the hospital, the professional association, and the regulator each hold a piece of the responsibility, and the absence of a clear owner means that the responsibility can fall between them.(26,36) This distribution is both a weakness and an opportunity: a weakness because it allows the responsibility to dissipate, and an opportunity because it means the standards can be shaped by the profession itself rather than imposed from outside, and the profession is better placed than any regulator to know what a dental model actually does in practice. The proposals for governance emphasize a shared responsibility in which the developer certifies the model, the clinician maintains oversight, and the institution audits the outcomes, but these roles are aspirational rather than established.(36)The distribution of responsibility is not merely a matter of assigning blame when something goes wrong; it is also a matter of who has the incentive and the authority to make the tool better, and a governance model that leaves no one clearly in charge also leaves no one clearly motivated to improve the tool.(31)

The international dimension adds a further complication. A model developed and validated in one jurisdiction may be deployed in another with a different regulatory regime, and the cross-border flow of both models and data means that the governance question cannot be answered entirely within a single legal framework.(5,27) A clinician in one country may use a tool certified in another without any assurance that the certification translates, and the same is true of the data on which the tool was trained, whose provenance may span jurisdictions with different privacy laws. The professional associations, which operate across jurisdictions, are therefore well placed to establish norms that outpace the slower process of legislation, and the guidelines they have begun to issue are an early form of the coordination that the field needs.(33,35) The trajectory of the debate, from permissive enthusiasm to a more guarded and structured approach, is itself worth noting, and the proposals that once treated artificial intelligence as a general-purpose tool are giving way to a recognition that a diagnostic model and a conversational assistant require different standards of scrutiny.(28,34)

Data, privacy and consent

The training of dental models depends on large collections of radiographs, photographs, and clinical records, and the governance of that data is the least resolved of the axes. The sources of training data are often unclear, the consent under which the data were originally collected did not anticipate reuse for machine learning, and the linkage between a training image and the patient it came from is not always severed.(37,38) The privacy risk is concrete, because a dental radiograph can identify a patient and can reveal more about them than their dental status alone.(39)The shift from a model of clinical data use, in which a radiograph serves a single patient, to a model of training data use, in which the same radiograph serves a population of future patients, is a change in kind that the existing governance does not fully capture.(38)The dental practice, which is often a small and lightly regulated setting compared with a hospital, is a particularly weak point in this chain, because the records it holds are rich in exactly the imaging data that the models need, and the governance of those records is frequently informal.

The tension between data sharing and privacy is structural. The models improve with more and more diverse data, which argues for sharing, while the privacy of the patients whose data are shared argues for restriction, and the reconciliation of the two is an open problem.(38,40) Techniques such as federated learning and de-identification mitigate the tension but do not eliminate it, and the consent model under which patients authorize the use of their data for training is still being worked out.(40,41) Federated learning, in which the model is trained across institutions without the data leaving each institution, addresses the privacy concern at the cost of a more complex and less transparent training process, and its adoption in dentistry is at an early stage.(40) De-identification, in which the data are stripped of identifying information, is more widely used but is not a complete solution, because a dental image itself can be identifying and because the demographic variables that fairness auditing requires are exactly the ones that de-identification removes.

The consent question is the sharpest point of this axis. A patient who consents to a radiograph for a clinical purpose has not thereby consented to its use in training a model, and the retrospective reuse of clinical data for training is governed by frameworks that were written before machine learning existed.(37)The dental community needs a consent model that is specific to machine learning, that distinguishes clinical use from training use, and that is honest about the fact that a model cannot be retracted from what it has learned.(37,41)

The provenance of the data is as important as the consent that governs it. A model is only as good as the data it was trained on, and if the provenance of that data is not documented, the model cannot be audited, cannot be held to a fairness standard, and cannot be reliably re-validated in a new setting.(37)The sources of training data, the criteria for inclusion and exclusion, and the demographic composition of the data are all part of the documentation that a responsible deployment requires, and their absence is itself a signal that the model has not been prepared for clinical use.(39,40)

Accountability and liability

The final axis is the allocation of responsibility when a model errs. When a clinician misreads a radiograph, the responsibility is clear; when a model misreads it and the clinician accepts the model's reading, the responsibility is distributed between the developer, the clinician, and the institution, and the law has not settled how to divide it.(42,43) The distinction between a tool that advises and a tool that decides is central to this question, because the law has well-established doctrines for human error but not yet for the mixed human-machine error that a decision support tool produces, and the novelty of the situation is what makes the responsibility hard to allocate.(42)The question is not academic, because a missed lesion or a wrong measurement that traces to a model will eventually reach a patient, and the question of who is accountable will then be litigated rather than debated.(42)The dental context adds a specific complication: many of the tools are marketed directly to practices as productivity aids rather than as medical devices, which may shift the perceived responsibility away from the developer and onto the clinician, even though the clinician has no control over the model's behavior.(43)

The concept of the human-in-the-loop is the prevailing answer, in which the clinician retains the final responsibility and the model is a decision support, but the concept presumes that the clinician can meaningfully override the model, which is not always the case when the model's reasoning is opaque.(21,44) The stronger the model, the more deference the clinician tends to give it, and the more the responsibility shifts toward the model without a corresponding legal shift, which is the accountability gap in its sharpest form.(44)The automation bias, the tendency to defer to an automated system even when its output conflicts with one's own observation, is well documented in other domains and is a foreseeable risk in dentistry, where the model's apparent precision can mask its actual uncertainty.(42)

The proposals for closing the gap emphasize documentation, audit, and shared responsibility. The model should be documented with its training data, its limitations, and its validation, the clinician should document the role the model played in each decision, and the institution should audit the outcomes, so that when an error occurs, its source can be traced.(43,45) These practices are feasible but not yet standard, and their adoption is the clearest near-term step the profession can take.

The question of liability interacts with the question of trust. A clinician who does not trust the model will override it even when it is right, and a clinician who trusts it too much will accept it even when it is wrong, and the calibration of that trust is a clinical skill that the current training does not teach.(32,44) The calibration is made harder by the fact that the model's confidence, where it is even reported, is not a reliable guide to its correctness, so the clinician cannot simply

defer to a high-confidence output or dismiss a low-confidence one.(44)The same concern applies to the patient, who may either reject a correct machine-assisted diagnosis out of distrust or accept an incorrect one out of deference to the machine. Building calibrated trust, in both clinicians and patients, is therefore as much a part of responsible deployment as the technical safeguards, and it is a task that the profession has only begun to consider.(22,32)

The relationship between trust and accountability is not a simple trade-off. A more transparent model may be easier to trust but harder to hold accountable, if the transparency exposes the complexity of the reasoning without exposing a clear locus of error, and a simpler model may be easier to hold accountable but less accurate. The choice between them is not purely technical but also ethical, because it weighs the value of an accurate answer against the value of an answer that a human can understand and answer for.(21)This is the deeper question beneath the accountability axis, and it is one the field has only begun to pose.

DISCUSSION

Across the five axes, a single pattern emerges. The accuracy literature has advanced far ahead of the ethical, bias-related, and regulatory infrastructure that a safe deployment requires, and the gap between the two is the defining characteristic of the field.(3,6) The models are good enough to be used, but the questions of whether they are fair, whether they are lawful, whether the data were consented, and whether anyone is accountable for their errors remain, for the most part, unanswered. This gap is not unique to dentistry; the broader medical field faces the same lag, and the dental community can learn from the responses that have been proposed there.(5,6,44)

The axes also interact. A biased model is an ethical failure, an unregulated model is a governance failure, and an undocumented model is an accountability failure, and the same tool can exhibit all three at once. The interaction means that the axes cannot be addressed one at a time; a tool that is made fair but remains unregulated is not yet safe, and a tool that is regulated but remains biased is not yet equitable.(5,20) The resolution must therefore be coordinated across the axes, and the institutions that can coordinate them are only beginning to form.(36)The concept of trustworthy artificial intelligence, which bundles fairness, transparency, and accountability into a single standard, is the most developed attempt at such coordination, and its adoption by the dental community would go a long way toward closing the gap.(20,21) The dental literature has begun to converge on this language, and the appearance of the term trustworthy in dental artificial intelligence publications is itself a sign that the field is moving from a purely technical to a more holistic frame.(20,26)

The limitations of this review are those of its sources and of its method. The dental literature on these axes is thin, and the review therefore leans on the broader medical literature and on conceptual arguments, which may not translate perfectly to the specific context of dentistry. The selection was narrative and not systematic, so the account is shaped by what was accessible and relevant rather than by an exhaustive search. The review also could not assess the rate at which the proposed practices are actually adopted, because that evidence does not yet exist. The landmark demonstrations of bias in medical algorithms are drawn from settings outside dentistry, and their direct relevance to dental models is plausible but not yet demonstrated.(4)These limitations should be read not as weaknesses of the review but as features of the field it describes: the evidence that a thorough account of these dimensions requires does not yet exist in dental form, and the review is necessarily as much an agenda as a summary.

The agenda that follows is concrete. For research, the field needs studies that report subgroup performance and fairness metrics alongside accuracy, and that document the provenance and consent of the training data. For regulation, the field needs a clear determination of which dental tools are devices and which are not, and a post-market mechanism for the models that update. For practice, the field needs a consent model that is specific to machine learning and a documentation standard that makes the model's role in each decision traceable.(3,27,43)

The agenda also has an educational component. Clinicians who will work alongside these tools need to understand what the tools can and cannot do, where their errors concentrate, and how to calibrate their trust, and this understanding is not yet part of the dental curriculum.(22,34) The same applies to patients, whose consent to a machine-assisted diagnosis is meaningful only if they understand what the machine contributed, and whose autonomy is served by an honest account of the tool's role rather than by a technical one. Education, in this sense, is not separate from the ethical and regulatory agenda; it is the mechanism by which the other measures become effective in practice.

The synthesis offered here is therefore a call for coordination rather than a set of isolated recommendations. The bias, ethics, regulation, data, and accountability axes each have their own literature and their own advocates, but a tool that satisfies one and neglects the others is not yet safe, and the patient is the one who bears the combined risk. The risk is not theoretical but personal, because it lands on a single patient at a single appointment, and the standards that protect that patient must be in place before the appointment, not discovered during it. The most useful single step the profession could take is to adopt a shared standard, modeled on the trustworthy artificial intelligence frameworks already proposed, that requires any dental model to demonstrate fairness, explainability, lawful status, data provenance, and a clear line of accountability before it is deployed.^(20,27) Whether such a standard is enforced by the regulator, the profession, or the institution matters less than that it exists and is applied.

The review closes with a note on pace. The technology is advancing faster than the frameworks that should govern it, and the distance between the two is widening rather than narrowing, because each new capability, from the first diagnostic models to the current generative systems, arrives before the previous one has been absorbed into practice.^(3,5) This is not an argument for slowing the technology, which would be neither feasible nor desirable, but for accelerating the governance, so that the questions raised here are answered in time to guide the adoption they concern rather than after the fact. The questions are already answerable in principle; what is missing is the will and the institution to answer them at the pace the technology demands. Building that will, and that institution, is the task to which this review ultimately points, and it is a task for the whole dental community, for its clinicians and its researchers alike, working together.

CONCLUSIONS

Artificial intelligence in dentistry has advanced in accuracy faster than in fairness, ethics, and regulation, and the safe adoption of the technology depends on closing that gap. Algorithmic bias is inherited from non-representative training data and is rarely audited; the classical ethical principles apply directly and require transparency and human oversight; the regulatory status of dental tools is unsettled; the governance of training data and consent is unresolved; and the allocation of responsibility for model errors remains unclear. These dimensions are connected, and their resolution requires coordinated action by developers, clinicians, institutions, and regulators. The clinical value of dental artificial intelligence will therefore depend as much on this infrastructure as on further gains in accuracy, and the two must advance together rather than in sequence. A tool that is accurate but biased, or fair but unregulated, or lawful but unaccountable, has not yet earned its place in clinical practice, and the standard by which these tools should be judged is the whole of these dimensions, not the accuracy metric alone.

REFERENCES

1. Liu TY, Lee KH, Mukundan A, Karmakar R, Dhiman H, Wang HC. AI in Dentistry: Innovations, Ethical Considerations, and Integration Barriers. *Bioengineering*. 2025;12(9):928. doi: 10.3390/bioengineering12090928
2. Wang T, Emami E, Jafarpour D, Tolentino R, Gore G, Rahimi SA. Integrating equity, diversity, and inclusion throughout the lifecycle of artificial intelligence for healthcare: a scoping review. *PLOS Digit Health*. 2025;4(7):e0000941. doi: 10.1371/journal.pdig.0000941
3. Schwendicke F, Sidhu SK, Ferracane JL, Tichy A, Jakubovics NS. Generative AI: Opportunities, Risks, and Responsibilities for Oral Sciences. *J Dent Res*. 2025;104(13):1429-1431. doi: 10.1177/00220345251356408
4. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366(6464):447-453. doi: 10.1126/science.aax2342
5. Giansanti D. The Regulation of Artificial Intelligence in Digital Radiology in the Scientific Literature: A Narrative Review of Reviews. *Healthcare*. 2022;10(10):1824. doi: 10.3390/healthcare10101824
6. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. *Nat Med*. 2022;28(1):31-38. doi: 10.1038/s41591-021-01614-0
7. 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
8. Alfaraj A, Nagai T, AlQallaf H, Lin WS. Race to the Moon or the Bottom? Applications, Performance, and Ethical Considerations of Artificial Intelligence in Prosthodontics and Implant Dentistry. *Dentistry Journal*. 2024;13(1):13. doi: 10.3390/dj13010013
9. Yang Y, Liu Y, Liu X, Gulhane A, Mastrodicasa D, Wu W, et al. Demographic bias of expert-level vision-language foundation models in medical imaging. *Sci Adv*. 2025;11(13):eadq0305. doi: 10.1126/sciadv.adq0305
10. Cirillo D, Catuara-Solarz S, Morey C, Guney E, Subirats L, Mellino S, et al. Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. *npj Digit Med*. 2020;3(1):81. doi: 10.1038/s41746-020-0288-5
11. Schuch HS, Furtado M, Silva GFDS, Kawachi I, Chiavegatto Filho ADP, Elani HW. Fairness of Machine Learning Algorithms for Predicting Foregone Preventive Dental Care for Adults. *JAMA Netw Open*. 2023;6(11):e2341625. doi: 10.1001/jamanetworkopen.2023.41625
12. Ueda D, Kakinuma T, Fujita S, Kamagata K, Fushimi Y, Ito R, et al. Fairness of artificial intelligence in healthcare: review and recommendations. *Jpn J Radiol*. 2024;42(1):3-15. doi: 10.1007/s11604-023-01474-3
13. Solanki P, Grundy J, Hussain W. Operationalising ethics in artificial intelligence for healthcare: a framework for AI developers. *AI Ethics*. 2023;3(1):223-240. doi: 10.1007/s43681-022-00195-z
14. Alderman JE, Palmer J, Laws E, McCradden MD, Ordish J, Ghassemi M, et al. Tackling algorithmic bias and promoting transparency in health datasets: the STANDING Together consensus recommendations. *The Lancet Digital Health*. 2025;7(1):e64-e88. doi: 10.1016/s2589-7500(24)00224-3
15. Sabir LB, Mohtashim F, Uddin SMF. Overcoming barriers to AI implementation in dentistry: a comprehensive ISM model analysis. *Int J Health Care Qual Assur*. 2025;1-18. doi: 10.1108/ijhcqa-04-2024-0042
16. Chen RJ, Wang JJ, Williamson DFK, Chen TY, Lipkova J, Lu MY, et al. Algorithmic fairness in artificial intelligence for medicine and healthcare. *Nat Biomed Eng*. 2023;7(6):719-742. doi: 10.1038/s41551-023-01056-8
17. Prakash S, Balaji JN, Joshi A, Surapaneni KM. Ethical Conundrums in the Application of Artificial Intelligence (AI) in Healthcare—A Scoping Review of Reviews. *JPM*. 2022;12(11):1914. doi: 10.3390/jpm12111914
18. Kumar P, Chauhan S, Awasthi LK. Artificial Intelligence in Healthcare: Review, Ethics, Trust Challenges & Future Research Directions. *Engineering Applications of Artificial Intelligence*. 2023;120:105894. doi: 10.1016/j.engappai.2023.105894
19. Arjomandi Rad A, Vardanyan R, Athanasiou T, Maessen J, Sardari Nia P. The ethical considerations of integrating artificial intelligence into surgery: a review. *Interdiscip Cardiovasc Thorac Surg*. 2025;40(3):ivae192. doi: 10.1093/icvts/ivae192
20. Ducret M, Wahal E, Gruson D, Amrani S, Richert R, Mouncif-Moungache M, et al. Trustworthy Artificial Intelligence in Dentistry: Learnings from the EU AI Act. *J Dent Res*. 2024;103(11):1051-1056. doi: 10.1177/00220345241271160
21. the Precise4Q consortium, Amann J, Blasimme A, Vayena E, Frey D, Madai VI. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Med Inform Decis Mak*. 2020;20(1):310. doi: 10.1186/s12911-020-01332-6
22. Di Mauro L, Capasso E, Tettamanti C, Casella C, Francaviglia M, Volonnino G, et al. The role of artificial intelligence in analyzing clinical malpractice disputes through medical record management. *Journal of Forensic and Legal Medicine*. 2025;115:102941. doi: 10.1016/j.jflm.2025.102941
23. Mohammad Amini M, Jesus M, Fanaei Sheikholeslami D, Alves P, Hassanzadeh Benam A, Hariri F. Artificial Intelligence Ethics and Challenges in Healthcare Applications: A Comprehensive Review in the Context of the European GDPR Mandate. *MAKE*. 2023;5(3):1023-1035. doi: 10.3390/make5030053
24. Arriola-Guillén LE. THE ROLE OF ARTIFICIAL INTELLIGENCE IN THE AUTHORSHIP OF SCIENTIFIC ARTICLES IN DENTISTRY: ETHICAL APPROACHES FOR RESPONSIBLE USE. *Rev Cient Odontol (Lima)*. 2025;14(1):e269. doi: 10.21142/2523-2754-1401-2026-269
25. Zhang J, Zhang ZM. Ethics and governance of trustworthy medical artificial intelligence. *BMC Med Inform Decis Mak*. 2023;23(1):7. doi: 10.1186/s12911-023-02103-9
26. Jha D, Durak G, Das A, Sanjotra J, Susladkar O, Sarkar S, et al. Ethical framework for responsible foundational models in medical imaging. *Front Med*. 2025;12:1544501. doi: 10.3389/fmed.2025.1544501
27. Ahmad M. Regulating intelligence: a systematic analysis of safety, ethics, and equity in artificial intelligence driven healthcare. *Intelligence-Based Medicine*. 2025;12:100320. doi: 10.1016/j.ibmed.2025.100320

28. Ratti E, Morrison M, Jakab I. Ethical and social considerations of applying artificial intelligence in healthcare—a two-pronged scoping review. *BMC Med Ethics*. 2025;26(1):68. doi: 10.1186/s12910-025-01198-1
29. Tang L, Li J, Fantus S. Medical artificial intelligence ethics: A systematic review of empirical studies. *DIGITAL HEALTH*. 2023;9:20552076231186064. doi: 10.1177/20552076231186064
30. Pathni RK. Artificial Intelligence and the Myth of Objectivity. *Journal of Healthcare Management Standards*. 2023;3(1):1-14. doi: 10.4018/jhms.329234
31. Stanley EAM, Souza R, Wilms M, Forkert ND. Where, why, and how is bias learned in medical image analysis models? A study of bias encoding within convolutional networks using synthetic data. *eBioMedicine*. 2025;111:105501. doi: 10.1016/j.ebiom.2024.105501
32. Stanley EAM, Souza R, Winder AJ, Gulve V, Amador K, Wilms M, et al. Towards objective and systematic evaluation of bias in artificial intelligence for medical imaging. *J Am Med Inform Assoc*. 2024;31(11):2613-2621. doi: 10.1093/jamia/ocae165
33. Chau M, Rahman MG, Debnath T. From black box to clarity: Strategies for effective AI informed consent in healthcare. *Artificial Intelligence in Medicine*. 2025;167:103169. doi: 10.1016/j.artmed.2025.103169
34. Corfmat M, Martineau JT, Régis C. High-reward, high-risk technologies? An ethical and legal account of AI development in healthcare. *BMC Med Ethics*. 2025;26(1):4. doi: 10.1186/s12910-024-01158-1
35. Goktas P, Grzybowski A. Shaping the Future of Healthcare: Ethical Clinical Challenges and Pathways to Trustworthy AI. *JCM*. 2025;14(5):1605. doi: 10.3390/jcm14051605
36. Allareddy V, Oubaidin M, Rampa S, Venugopalan SR, Elnagar MH, Yadav S, et al. Call for algorithmic fairness to mitigate amplification of racial biases in artificial intelligence models used in orthodontics and craniofacial health. *Orthod Craniofac Res*. 2023;26(S1):124-130. doi: 10.1111/ocr.12721
37. Celi LA, Cellini J, Charpignon ML, Dee EC, Dernoncourt F, Eber R, et al. Sources of bias in artificial intelligence that perpetuate healthcare disparities—A global review. *PLOS Digit Health*. 2022;1(3):e0000022. doi: 10.1371/journal.pdig.0000022
38. Siala H, Wang Y. SHIFTing artificial intelligence to be responsible in healthcare: A systematic review. *Social Science & Medicine*. 2022;296:114782. doi: 10.1016/j.socscimed.2022.114782
39. Karimian G, Petelos E, Evers SMAA. The ethical issues of the application of artificial intelligence in healthcare: a systematic scoping review. *AI Ethics*. 2022;2(4):539-551. doi: 10.1007/s43681-021-00131-7
40. Drukker K, Chen W, Gichoya JW, Grusauskas NP, Kalpathy-Cramer J, Koyejo S, et al. Toward fairness in artificial intelligence for medical image analysis: identification and mitigation of potential biases in the roadmap from data collection to model deployment. *J Med Imag*. 2023;10(06):1-26. doi: 10.1117/1.jmi.10.6.061104
41. Nwebonyi N, McKay F. Exploring bias risks in artificial intelligence and targeted medicines manufacturing. *BMC Med Ethics*. 2024;25(1):113. doi: 10.1186/s12910-024-01112-1
42. Murphy K, Di Ruggiero E, Upshur R, Willison DJ, Malhotra N, Cai JC, et al. Artificial intelligence for good health: a scoping review of the ethics literature. *BMC Med Ethics*. 2021;22(1):14. doi: 10.1186/s12910-021-00577-8
43. Arora A, Alderman JE, Palmer J, Ganapathi S, Laws E, McCradden MD, et al. The value of standards for health datasets in artificial intelligence-based applications. *Nat Med*. 2023;29(11):2929-2938. doi: 10.1038/s41591-023-02608-w
44. Kale AU, Dattani R, Tabansi A, Hogg HDJ, Pearson R, Glocker B, et al. AI as a Medical Device Adverse Event Reporting in Regulatory Databases: Protocol for a Systematic Review. *JMIR Res Protoc*. 2024;13:e48156. doi: 10.2196/48156
45. Dettori M, Lamloum D, Lingström P, Campus G. Artificial Intelligence and Innovation in Oral Health Care Sciences: A Conceptual Review. *Healthcare*. 2025;13(24):3327. doi: 10.3390/healthcare13243327

FUNDING

No financing

CONFLICT OF INTEREST

None

AUTHORSHIP CONTRIBUTIONS

Drafting – original draft: Youris Samón Mayan, Marta María Capote Fernández.

Writing–review and editing: Youris Samón Mayan, Marta María Capote Fernández.