REVIEW

Received / Geliş Tarihi: 01.07.2026
Accepted / Kabul Tarihi: 13.08.2026
Published / Yayın Tarihi: 31.08.2026

Aydın Dental Journal - Volume **12** Issue **2** - August **2026** (147-159)

Aydın Dental Journal

Journal homepage: http://dergipark.ulakbim.gov.tr/adj

DOI: 10.17932/IAU.DENTAL.2015.009/dental_v012i2006

Artificial Intelligence and Dentistry: A Clinical Review and a Proposed CHAIR Framework

Yapay Zekâ ve Diş Hekimliği: Bir Klinik Derleme ve Çerçeve Önerisi (CHAIR)

Doğan Şengül*

ABSTRACT

Objective: Artificial intelligence (*AI*) is entering dental practice through imaging software, screening initiatives and patient-facing chatbots, yet clinicians have limited concise guidance on how to evaluate these tools. This review summarises where *AI* now meets dentistry and offers a practical method for appraising a tool before chairside use.

Methods: We conducted a narrative review of peer-reviewed studies, regulatory texts and reporting guidelines, favouring work that was validated beyond its own development sample, deployed in real settings or explicit about bias and governance. The evidence is organised along the clinical pathway and paired with a proposed five-point checklist (the *CHAIR* framework) and a worked example.

Results: The strongest evidence lies in radiographic image analysis. Convolutional neural networks detect caries on bitewings, measure periodontal and periapical bone loss and segment teeth and bone on cone-beam computed tomography at a level comparable with experienced clinicians. Photograph-based models screen for oral cancer and its precursor lesions, while other systems assist cephalometric analysis, extraction planning and implant identification. Language models are starting to help with documentation and patient communication. Accuracy is highest on tightly defined tasks and usually declines when the scanner, patient group or disease distribution differs from the system's training data; reliable testing in new clinical settings remains uncommon.

Conclusion: AI is most useful when framed as augmented intelligence: a support for the dentist's judgement, not a substitute. Safe adoption depends on appraising each tool against the local setting, monitoring for bias and putting clear governance in place.

Keywords: *Artificial intelligence, Clinical decision support systems, Dentistry, Risk management, Workflow integration.*

ÖZET

Amaç: Yapay zekâ (*YZ*), görüntüleme yazılımları, tarama girişimleri ve hastaya yönelik sohbet robotları aracılığıyla diş hekimliği pratiğine girmektedir. Diğer yandan, klinisyenlerin bu araçları nasıl değerlendireceğine ilişkin kısa ve uygulanabilir rehberlik hâlen sınırlıdır. Bu derleme, *YZ*'nin diş hekimliğiyle güncel kesişim alanlarını özetlemekte ve bir aracın klinik kullanım öncesinde değerlendirilmesi için pratik bir yöntem sunmaktadır.

Gereç ve Yöntemler: Hakemli çalışmalar, düzenleyici metinler ve raporlama kılavuzları temel alınarak anlatısal bir derleme yapılmıştır. Kendi geliştirme veri kümesi dışında doğrulanmış, prospektif olarak değerlendirilmiş, gerçek klinik ortamlarda uygulanmış ya da yanlılık ve yönetişim konularını açıkça ele almış çalışmalara öncelik verilmiştir. Kanıtlar klinik bakım akışı boyunca düzenlenmiş; önerilen beş maddelik bir kontrol listesi (*CHAIR* çerçevesi) ve uygulamalı bir örnekle birlikte sunulmuştur.

Bulgular: En güçlü kanıtlar radyografik görüntü analizinde yoğunlaşmaktadır. Evrişimli sinir ağları, bite-wing radyografilerinde çürük lezyonlarını saptamakta, periodontal ve periapikal kemik kaybını ölçmekte ve konik ışınlı bilgisayarlı tomografide diş ve kemik dokularının segmentasyonunu deneyimli hekimlerle karşılaştırılabilir doğrulukta gerçekleştirmektedir. Fotoğraf tabanlı modeller ağız kanseri ve öncül lezyonların taranmasına katkı sağlarken, diğer sistemler sefalometrik analiz, çekim planlaması ve implant tanımlamayı desteklemektedir. Dil modelleri klinik dokümantasyon ve hasta iletişiminde yardımcı olmaya başlamıştır. Doğruluğun dar tanımlı görevlerde en yüksek olduğu; tarayıcı, hasta grubu veya hastalık dağılımı sistemin eğitim verilerinden uzaklaştığında genellikle düştüğü anlatılmaktadır. Yeni bir klinik ortamda güvenilir doğrulama uygulamaları ise hâlen nadirdir.

Sonuç: *YZ* en yararlı biçimde, diş hekiminin kararını destekleyen ancak onun yerini almayan artırılmış zekâ olarak değerlendirilmelidir. Güvenli kullanım, her aracın yerel koşullara göre değerlendirilmesine, yanlılık açısından izlenmesine ve açık yönetişim mekanizmalarının kurulmasına bağlıdır.

Anahtar Kelimeler: *Diş hekimliği, İş akışı entegrasyonu, Risk yönetimi, Sağlık teknolojilerinin benimsenmesi, Yapay zekâ.*

[1]Assistant Professor, Department of Software Engineering, Faculty of Engineering and Natural Sciences, İstanbul Sabahattin Zaim University, İstanbul, Türkiye

***Corresponding Author:** Doğan Şengül, **e-mail:** dogan.sengul@izu.edu.tr, **ORCID:** 0000-0002-2285-3907, Department of Software Engineering, Faculty of Engineering and Natural Sciences, İstanbul Sabahattin Zaim University, Halkalı, Küçükçekmece, İstanbul, Türkiye

Introduction

In recent years, artificial intelligence (AI) has transitioned from experimental research into active clinical practice. The medical field no longer views AI as a novel concept, but rather as a routine tool for diagnosis, documentation, and patient triage.[1] The dental profession is also actively participating in this technological shift. A practitioner may open a new version of an imaging package and find that caries are already outlined on the bitewing, receive a panoramic report in which every tooth has been numbered automatically or meet a patient who has rehearsed a diagnosis with a chatbot before sitting down. The question for the profession is no longer whether these tools will arrive, but how they should be evaluated and integrated safely.

Several features of dental practice make it unusually receptive to image-based computation. Much of the diagnostic day is spent reading standardised pictures, the bitewing, the periapical film, the panoramic radiograph and, increasingly, the cone-beam computed tomography (*CBCT*) volume, and these are produced in very large numbers and stored digitally.[2] Many dental images contain recurring visual patterns and measurable structures, such as interproximal lesions and crestal bone levels, that are well suited to computational support. It is therefore not surprising that some of the earliest and most convincing demonstrations of clinical *AI* in dentistry have concerned the radiograph.

This article is written for dental clinicians evaluating AI claims in practice, rather than for engineers developing the models. It does not set out to be the first survey of the field: broad overviews have already charted the spread of *AI* across the discipline,[2] systematic reviews have catalogued its applications and measured its accuracy,[3] and others have examined the ethical questions it raises.[4] The aim here is narrower and more practical. After a short glossary, the review follows the patient through the clinic, from the first radiograph to the conversation that closes the visit, and shows where the evidence is strongest and where it remains preliminary. It then turns to the weaknesses that recur across the field, sets out a proposed five-point method for appraising a tool before adoption (the *CHAIR* framework) and works through a concrete example. Throughout, the regulatory setting of Türkiye and the European Union is kept in view, because the safe use of these tools is as much a matter of governance as of accuracy. This review is not intended only as a broad survey of artificial intelligence in dentistry; its main contribution is a practical CHAIR checklist that helps dentists evaluate whether an AI tool is clinically relevant, locally valid, safe and governable before chairside use.

Scope And Literature Search

This is a narrative rather than a systematic review, and it makes no claim to completeness. Relevant work was identified in PubMed and Scopus and supplemented by following the reference lists of major reviews and by consulting regulatory and reporting-guideline sources. The queries paired artificial intelligence, machine learning, deep learning and large language models with dentistry, dental radiography, caries, periodontitis, cone-beam computed tomography, oral cancer, orthodontics, endodontics and dental implants. Studies were favoured when they were checked beyond the data on which they were built, whether externally or prospectively, when they reported use in real clinics or when they confronted questions of bias, transparency or governance directly; recent systematic reviews were also drawn upon. The final literature search was completed in June 2026. No starting-date restriction was applied, although the cited evidence concentrates on work published from 2016 onwards, and only publications in English were considered. Titles and abstracts were screened for clinical relevance to dentistry and full texts were then read against three informal criteria: a clearly described dental task, a stated reference standard and reported diagnostic or predictive performance. Where several studies addressed the same task, preference in the text was given to those that were externally or prospectively validated, multicentre or the largest of their kind and to systematic reviews and meta-analyses over single-centre reports. Because the synthesis is narrative, no formal selection diagram or risk-of-bias assessment is presented.

As the review rests only on published work and involved no patients, human participants or animals, ethics committee approval was not required.

Terminology And Core Concepts

A handful of terms recur throughout the literature, and a clear picture of how they fit together makes the rest of this review easier to read (Table 1, Figure 1). Artificial intelligence is the broad umbrella: software that performs tasks usually thought to require human judgement, such as recognising a pattern, interpreting language or supporting a decision.[1] Nested inside it is machine learning (*ML*), the family of methods that get better at a job by drawing patterns from examples rather than from rules a programmer writes by hand. Deeper still sits deep learning (*DL*), built on artificial neural networks with many layers.

Where images are concerned, the dominant tool is the convolutional neural network (*CNN*); it sits beneath almost every radiograph-based and photograph-based system discussed here. For words, a different design, the transformer, powers the large language models (*LLM*) now entering clinics, while natural language processing (*NLP*) is the wider field concerned with reading and generating human language[2]. Two further ideas matter at the chairside. Generative *AI* produces new content, whether a draft referral letter or the plain-language answer a patient receives from a chatbot. Foundation models are very large networks first trained on huge unlabelled collections and then tuned to a particular task with comparatively few labelled examples, which is appealing wherever expert annotation is hard to obtain.[5] Two cautions complete the picture. What is marketed as *AI* in medicine is better described as augmented intelligence, since the software is there to support the clinician rather than to take their place; and these systems are narrow, so a model that excels at grading caries knows nothing of periodontal bone, and strong performance on one task or in one clinic does not transfer automatically to another.

Table 1. A chairside glossary of artificial-intelligence terms used in this review.

Term	What it means	Example in dentistry
Artificial intelligence (*AI*)	Software performing tasks that usually need human judgement	Outlining caries on a bitewing
Machine learning (*ML*)	Methods that learn from examples rather than fixed rules	A model improving as more films are labelled
Deep learning (*DL*)	*ML* built on many-layered neural networks	Most current dental image systems
Convolutional neural network (*CNN*)	A network designed to read images	Detecting periapical lesions on *CBCT*
Natural language processing (*NLP*)	Methods that read and generate language	Structuring a free-text clinical note
Large language model (*LLM*)	A network trained on vast text that produces fluent answers	Drafting a referral or a patient leaflet
Generative *AI*	Systems that create new content	A chatbot answering a patient's question
Foundation model	A large pretrained model adaptable with few labels	Re-using a general image model for a niche task
Narrow *AI*	A model good at a single task only	A caries model that cannot grade bone
Augmented intelligence	*AI* that supports rather than replaces the clinician	A second-reader overlay on a radiograph
Receiver operating characteristic (ROC) curve	A plot of sensitivity against 1 − specificity across all decision thresholds	Comparing two caries detectors independently of any single threshold
Area under the curve (AUC)	A single figure summarising the ROC curve, from 0.5 (chance) to 1.0 (perfect discrimination)	A supplier reporting an AUC of 0.92 as its headline discrimination measure.
Calibration	Agreement between predicted probabilities and observed frequencies	Lesions flagged with 80% confidence proving real about eight times in ten
Internal validation	Testing on data held back from the same source used to build the model	A random split of one clinic's bitewings
External validation	Testing on independent data from other patients, centres or devices	Trying a model on another clinic's radiographs before purchase

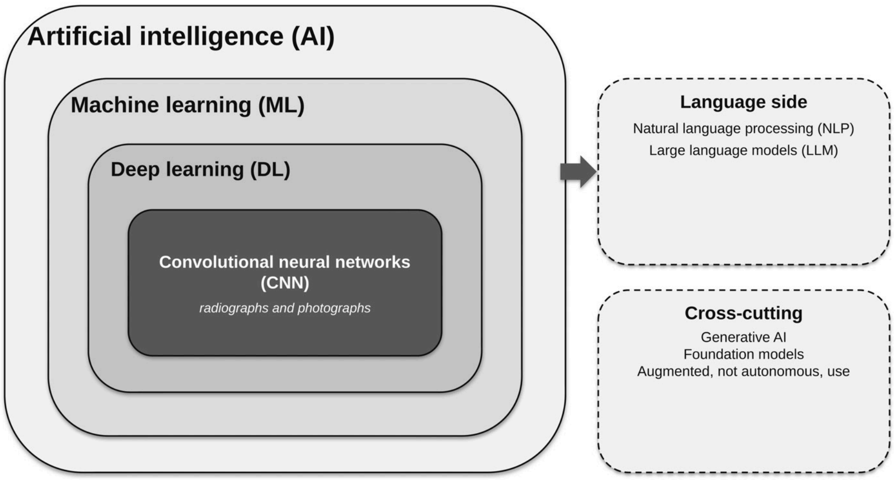

Figure 1. Nested families of artificial intelligence relevant to dentistry. Convolutional neural networks read images; a separate, language-oriented branch underlies large language models, and several ideas cut across both.

Current Applications of Artificial Intelligence in Dentistry

The clearest way to see what *AI* can and cannot yet do is to follow it along the clinical pathway, from the routine radiograph to the soft-tissue examination and the conversation that closes the visit (Table 2).

Caries Detection On Bitewing Radiographs

Caries detection on intraoral radiographs is the most thoroughly studied application in the field, and the results are consistent. Trained on thousands of annotated bitewings, a convolutional neural network reported higher sensitivity than independent dentists in that study population, and was particularly strong on the early, enamel-confined lesions that may be missed during routine interpretation.[6] Rather than replace the clinician, such a model is most useful as a second reader: when its output was shown to dentists, their own sensitivity rose without a collapse in specificity.[7] Systems developed on local data, including a Turkish model that both detects and outlines lesions, report comparable performance and show that the capability is not confined to a few large centres.[8] A meta-analysis confined to bitewing radiographs placed pooled sensitivity and specificity in a clinically useful range but again stressed how rarely these systems are tested outside their development data.[9]

Periodontal And Periapical Bone Assessment

The same modelling approach has been applied to the supporting structures of the tooth. A deep network trained to measure radiographic bone loss on panoramic images estimated the proportion of bone lost close to the assessment of experienced examiners, offering a route to faster and more consistent periodontal charting.[10] Building on this, hybrid systems have gone further and assigned a periodontitis stage from the radiograph, a step towards automated severity grading.[11] Systematic reviews and meta-analyses of these tools report good pooled accuracy for detecting bone loss and staging periodontitis, while noting wide variation between datasets and a continuing need for better reporting.[12] At the apex, a *CNN* applied to *CBCT* volumes detected periapical lesions with high reliability and estimated their size, a task in which small lesions are easily overlooked on two-dimensional films.[13] In each case the appeal is the same: a quantitative, reproducible assessment of structures that are currently evaluated visually, with recognised inter-observer variability.

Panoramic and Cone-Beam Imaging

Three-dimensional imaging is laborious to interpret, and this is where automation promises to save the most time. A large multicentre system segmented individual teeth and the surrounding alveolar bone from *CBCT* scans fully automatically and, for the teeth, performed at a level comparable with expert radiologists, producing the kind of three-dimensional model on which digital implant and orthodontic planning depend.[14] On panoramic radiographs, networks reliably detect and number teeth, a deceptively simple task that underpins automated charting and record-keeping.[15] Others map the relationship between an impacted lower third molar and the inferior alveolar nerve, the anatomical question that governs the risk of surgery in that region.[16] Even a notoriously difficult diagnosis, the vertical root fracture, has been approached with deep learning on panoramic images, though performance here is more modest and underlines the limits of the plain film.[17]

Oral Cancer and Soft-Tissue Screening

Beyond hard-tissue imaging, photograph-based models address a problem of real public-health weight: the late detection of oral cancer. A deep learning algorithm trained on tens of thousands of clinical photographs distinguished oral cavity squamous cell carcinoma (*OCSCC*) from normal mucosa with very high discrimination, raising the prospect of triage where a specialist is not immediately available.[18] Recognising that the greater clinical value lies upstream, other groups have classified and localised oral potentially malignant disorders (*OPMD*), the lesions that precede malignancy, again from ordinary photographs.[19] The promise is a low-cost aid to opportunistic screening; the caution, returned to below, is that a photograph captured on an unfamiliar device or in unfamiliar light may not resemble the images on which the model was trained.

Orthodontics, Endodontics and Prosthodontics

AI has also reached the planning side of dentistry. In orthodontics, automated cephalometric analysis locates the landmarks on which diagnosis depends, and a deep-learning system identified them with high reproducibility compared with repeat human tracings, which are known to show some intra-examiner variability.[20] An umbrella review of systematic reviews is more cautious, noting that automated landmarking still exceeds the accepted error for many points and continues to need an experienced orthodontist's oversight.[21] Machine-learning models trained on treated cases have gone on to predict the extraction decision, the central judgement of orthodontic planning, with accuracy approaching that of clinicians.[22] Endodontics has drawn comparable interest, and reviews of the field describe *AI* applied to periapical diagnosis, root canal anatomy, working-length determination and the prediction of treatment outcomes.[23] Elsewhere, networks recognise the brand and type of a dental implant from a radiograph, a small but genuinely useful problem when a patient presents years later with no record of the fixture they carry.[24] These examples share a theme: the algorithm handles a narrow, well-defined sub-task and hands the result back to the clinician, who keeps the wider decision.

Language Models in Clinical Practice

Recent systems increasingly work with text rather than images. Large language models can draft notes, summarise a history or answer a patient's question in plain language, and a scoping review finds them already applied across dental specialties, most often to patient queries and education though usually through a single general-purpose model.[25] In practice, the systems patients and clinicians most often encounter are general-purpose chatbots such as ChatGPT (OpenAI), Gemini (Google), Claude (Anthropic) and Copilot (Microsoft), rather than tools built for dentistry. Reported performance on dental knowledge tasks has improved in recent studies, and the choice of program matters: in the first head-to-head comparison of large language models on United States Integrated National Board Dental Examination questions, ChatGPT-4 answered 75.88% of items correctly, ahead of Claude-2.1 (66.38%) and Mistral-Medium (54.77%), although they remain uneven across topics and can be confident but incorrect.[26] Direct comparisons of these chatbots on diagnostic tasks, as distinct from examination-style knowledge, remain scarce in dentistry, and none of the general-purpose systems is authorised as a diagnostic medical device, so their output is best treated as text to be verified rather than as clinical advice. Because many patients now consult such a tool before the appointment, the dentist increasingly meets patients who have encountered information of variable reliability before the appointment. Looking further ahead, multimodal models that pair image with text could one day read a radiograph and explain it in plain language, though they sharpen rather than settle the questions of data privacy and bias.[27] The sensible posture is to treat these systems as a support for communication and paperwork, always checked by the clinician, and never as an unsupervised source of clinical advice, for reasons taken up in the next section.

Table 2. Representative applications of artificial intelligence across the dental care pathway.

Clinical area	Typical task	Illustrative evidence	Main caveat
Caries (bitewing)	Detect enamel and dentine lesions	More sensitive than dentists for early lesions; useful as a second reader	Hardest on incipient lesions; small datasets
Periodontal and periapical bone	Measure bone loss; stage disease; find apical lesions	Bone-loss and staging close to experts; reliable apical detection on *CBCT*	Definitions and reference standards vary
3D and panoramic imaging	Segment teeth and bone; number teeth; map the nerve	Multicentre *CBCT* segmentation matching expert radiologists	Heavy computation; sensitive to scanner
Oral cancer and *OPMD*	Screen photographs for malignant and pre-malignant lesions	High discrimination for carcinoma and precursor lesions	Lighting and camera differences; limited external testing
Orthodontics	Locate cephalometric landmarks; predict extraction	Highly reproducible landmarks; extraction decisions near clinicians	Trained on selected, treated cases
Endodontics	Detect vertical root fracture; assess canals	Workable detection on panoramic films	Modest on the plain film; stronger on *CBCT*
Implants and prosthetics	Identify implant brand and type	Accurate recognition from radiographs	Limited to systems seen in training
Documentation and communication	Draft notes; answer patient questions	Improving accuracy on board-style questions	Can be fluent but incorrect; needs checking

Across these applications, the quality of the underlying evidence deserves as much attention as the headline accuracies. Most published studies are retrospective, test their models on data drawn from the same source as the training material (internal validation) and rely on reference standards that vary from study to study; external validation on independent patients, devices or centres remains the exception, and prospective evaluation in routine care is rarer still.[9,28] Where AI has been compared with clinicians, the comparisons have often rested on bias-prone designs and small comparison groups[29] and even in the best-studied task, caries detection on bitewings, the pooled evidence comes from studies of heterogeneous design and quality.[9] Published accuracy therefore cannot be taken at face value: whether a result transfers depends on how closely the study population, equipment and workflow resemble those of the adopting clinic[30] and structuring that judgement, rather than reproducing any single reported figure, is the purpose of the appraisal framework proposed later in this review.

Barriers To Clinical Translation

Generalisability And Domain Shift

A central consideration for any clinical algorithm is that it learns the data it is shown, with all their particularities. A model trained on one scanner, one population or one mix of disease often performs worse when any of these changes, a failure known as domain shift (Figure 2). The pattern is well documented across medical imaging, where meta-analyses report excellent average accuracy alongside wide variation and frequent methodological weakness,[28] and where careful comparisons find the gap between algorithm and clinician to be smaller, and less consistent, than individual papers suggest.[29] The practical implication is clear: a published accuracy figure describes the test set on which it was measured, not the clinic about to adopt the tool, and the evidence that truly matters is testing carried out in patients and on devices that resemble the adopting practice's own.[30]

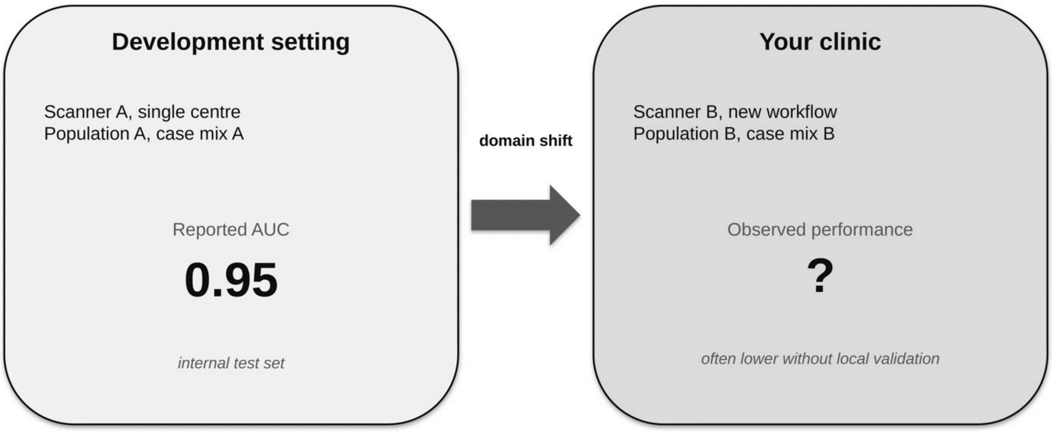

Figure 2. Why performance does not always transfer. A model can score highly on its own internal test set yet perform differently in another clinic with a different scanner, population and case mix, an effect known as domain shift.

Bias and Fairness

Because algorithms learn from their training data, they easily copy existing inequalities.[31] For example, deep neural networks can detect a patient's race from medical images without obvious features, unintentionally using demographic data.[32] Consequently, models trained on specific groups, such as adults or certain skin tones, often fail on others. Rather than rejecting AI, developers must provide clear data reporting. Clinics must test software across diverse patient groups before adoption and monitor it during daily use. Ultimately, institutions must actively measure fairness instead of assuming neutrality.

Model Opacity and Hallucination

Advanced AI systems often lack transparency. While "explainable *AI*" tools like heat maps provide visual clues, they do not guarantee correct logic; therefore, clinicians must never let AI override their own findings.[33] Additionally, language models frequently generate false information and fake references. Consequently, staff must manually verify all *AI*-generated facts, and clinics must strictly prohibit these systems from making independent patient decisions.

Privacy and Data Protection

Because dental images contain detailed, frequently shared personal data, strict protection is essential. In the *EU*, institutions must comply with the General Data Protection Regulation (*GDPR*) and the Artificial Intelligence Act, which classifies medical *AI* as high-risk and mandates rigorous safety measures.[34] Similarly, clinics in Türkiye must follow the Personal Data Protection Law (*KVKK*).[35] These rules are especially relevant in two situations: with children, for whom long-term data reuse raises special consent and stewardship concerns, and with any plan that moves images onto an external provider's servers. Because the dentist holds these records in trust, the handling of data should be spelled out plainly in the consent obtained for any service or study that involves *AI*, a duty underscored by reviews of the ethics of dental *AI*.[4]

Regulation and Reporting Standards

A growing number of *AI*-enabled devices have been authorised by regulators, including the United States Food and Drug Administration (*FDA*), yet authorisation is not the same as proof of patient benefit, and relatively few systems have been tested prospectively against clinical outcomes.[36] Reporting is a recurring weakness. When *AI* has been compared with clinicians systematically, the studies have tended to carry a high risk of bias, to lean on small comparison groups and to lack testing beyond the data on which the models were built,[29] which is why reporting standards such as the CONSORT-*AI* extension, with its companion guidelines for trial protocols and diagnostic-accuracy studies, were developed.[37] Related checklists extend the same discipline to the studies dentists most often meet, the Checklist for Artificial Intelligence in Medical Imaging for diagnostic-imaging work and the TRIPOD+*AI* statement for prediction models.[38,39] Familiarity with these standards supports critical appraisal of AI studies and marketing claims without requiring specialist computer-science training and helps clinicians recognise when the evidence behind a summary performance metric is limited.

The CHAIR Framework: A Five-Point Appraisal of Dental AI Tools

A new algorithm warrants the same care a clinician would give to any unfamiliar diagnostic test, and a short, memorable checklist turns that care into a habit. The five checks below can be recalled through the word *CHAIR*, and they follow a tool through its life, from the moment it is chosen to its long-term monitoring (Table 3, Figure 4).

CHAIR also occupies a different position from the reporting guidelines introduced above. CLAIM, CONSORT-AI and TRIPOD+AI are written primarily for researchers, reviewers and editors: across several dozen items each, they specify how a diagnostic-imaging study, a clinical trial or a prediction model should be reported, and they presume access to the full study report.[37-39] The dentist deciding whether to adopt a tool needs something shorter that answers a different question: not whether a study is well reported, but whether this tool should be used on these patients, with this equipment, in this workflow. The five domains of CHAIR were therefore distilled from the recurring failure points documented in the preceding section, mismatch between a tool's intended context and the clinic (C), undisclosed data and validation (H), opaque or single-figure accuracy claims (A), unresolved responsibility at integration (I) and absent monitoring after deployment (R), so that each check turns a documented weakness of the field into a question a clinician can put to a supplier. CHAIR is intended to complement rather than replace the reporting standards: a supplier able to answer its five questions should also be able to point to CLAIM- or TRIPOD+AI-compliant evidence, and a tool evaluated and transparently reported in accordance with CONSORT-AI should provide much of the information needed to assess the H and A checks.

Clinical question and context. Start from the decision the tool says it will sharpen, the patients and teeth it targets and the exact scanner, sensor or camera it assumes. A model proven on one manufacturer's bitewings is not automatically fit for another's.

Have the data and validation been shown? Ask how many images, from how many patients and centres, lie behind the tool, and whether their age range, case mix and equipment resemble the practice that will use it. The decisive question is whether the system has been tested externally or prospectively in a comparable setting, rather than only on a slice of its own development data. In the vocabulary of the field, testing on data held back from the tool's own development material is internal validation, and testing on independent data from other patients, centres or devices is external validation (Table 1).

Accuracy reported transparently. A single summary number, such as the area under the curve (*AUC*), does not provide enough clinical information. Developers must report sensitivity and specificity separately because these metrics trade off: maximizing detection increases false positives, while high specificity misses actual lesions (Figure 3). Evaluating accuracy across specific patient subgroups is equally crucial. The AUC is the area under the receiver operating characteristic (*ROC*) curve and summarises discrimination across all possible decision thresholds, from 0.5 (chance) to 1.0 (perfect); it does not show the trade-off actually chosen in use, which is why the paired measures matter. Where a tool outputs probabilities, their calibration, the agreement between predicted probabilities and the frequencies observed in reality, should also be reported (Table 1).

Integration and responsibility. Clinics must define who receives and has the authority to override *AI* results. If a system causes excessive false alarms or confuses medical responsibility, its disadvantages will outweigh its clinical benefits.

Risk monitoring after go-live. After implementing the software, institutions must actively track its performance. This requires assigning staff to regularly audit its accuracy and fairness across different patient populations. Facilities must also watch for any drop in performance when patient demographics or medical equipment change over time. If a developer does not provide clear guidelines for these ongoing checks, their tool is not ready for routine clinical use.

Diagnostic performance should be reported as sensitivity and specificity rather than as a single summary figure, because the two move in opposite directions as the operating threshold changes (Figure 3).

In practice, CHAIR is intended as a structured appraisal and a documentation habit rather than a scoring instrument. The five questions can be worked through with the supplier's evidence in a single sitting, the answers recorded briefly and kept, and the appraisal repeated whenever the tool, the imaging equipment or the patient profile changes materially. The checks do not carry equal weight. C, H and A concern the evidence itself and should be regarded as essential before adoption: a tool whose intended context does not match the clinic, whose data and validation cannot be shown or whose accuracy is not reported transparently should not be adopted, whatever its other attractions. I and R concern the conditions of use rather than the evidence: they do not rule a tool out in advance, but adoption should wait until the roles, override authority and audit schedule they require are in place. A tool that fails any one of the essential checks is better deferred than adopted provisionally.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \qquad \text{Specificity} = \frac{TN}{TN + FP}$$

Figure 3. Sensitivity and specificity, the paired measures that a single accuracy figure conceals. TP, true positives; FN, false negatives; TN, true negatives; FP, false positives.

Table 3. The *CHAIR* framework for appraising a dental artificial-intelligence tool.

Check	Ask	What good looks like	Red flag
C: Clinical question and context	Which decision, patients, teeth and device?	A defined task on the equipment you actually use	Works for caries, with no specifics
H: Have the data and validation?	How many images, from where and tested how?	External or prospective testing in a similar population	Only internal test-set results
A: Accuracy reported transparently	Sensitivity, specificity, calibration and subgroups?	Separate metrics by lesion type and subgroup	A single *AUC* headline
I: Integration and responsibility	Who sees, who overrides and what alert burden?	Clear roles and a manageable workload	Unclear liability; constant alerts
R: Risk monitoring after go-live	Who audits drift and fairness and when?	A named owner and a fixed audit schedule	No plan once the tool is live

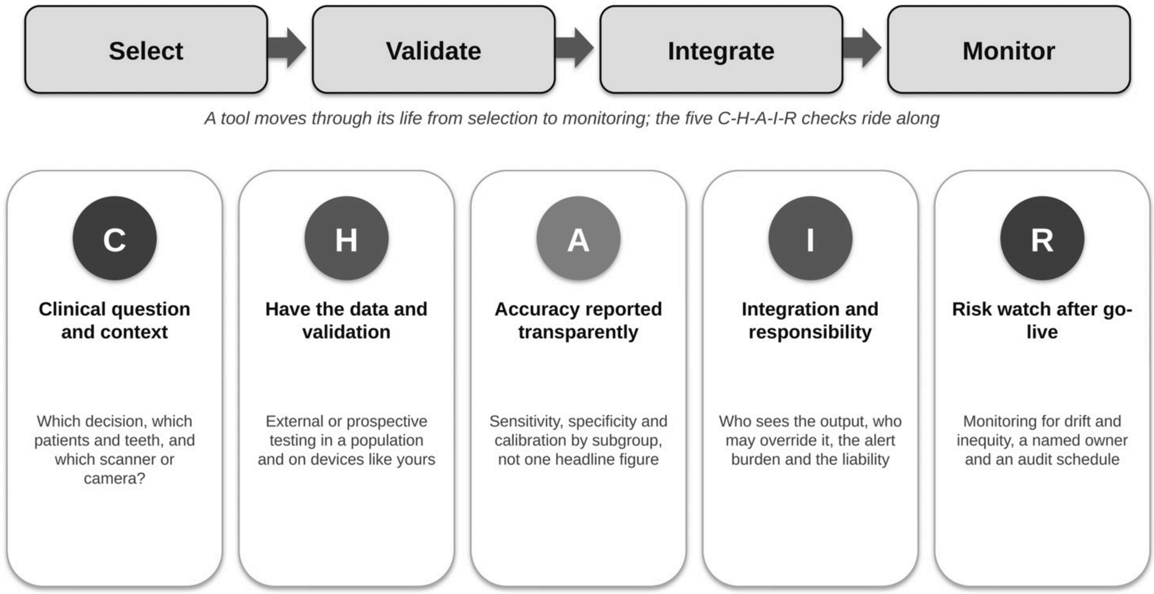

Figure 4. The *CHAIR* framework across the life of a dental *AI* tool. The five checks accompany a tool as it is selected, validated, integrated into the workflow and monitored after deployment.

CHAIR In Practice: A Worked Example in Bitewing Caries Detection

When a supplier offers an *AI* decay-detection tool claiming human-level accuracy (e.g., *AUC* 0.92), clinics should evaluate it using the *CHAIR* framework (Figure 5). The team must first confirm if the tool works on their specific patient demographic and X-ray sensors. Next, they must verify the training data and request sensitivity and specificity rates categorized by lesion depth. For workflow integration, staff must establish who reviews the *AI*'s results and how to handle clinical disagreements. Finally, the clinic must schedule performance monitoring at six and twelve months. If the supplier only provides the basic *AUC* score, the clinic should reject the tool. This evaluation process applies to any clinical *AI* software.

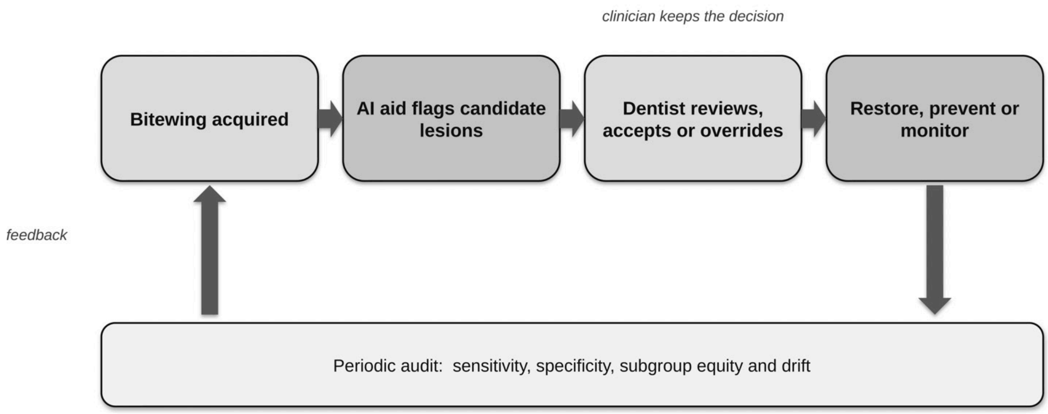

Figure 5. Where an artificial-intelligence aid sits in the bitewing workflow. The software flags candidate lesions, but the clinician keeps the decision, and a periodic audit feeds back into safe use.

Implications for the Profession: Education, Governance and the Turkish Context

While dental professionals show strong interest in AI, many lack formal evaluation training. To address this, universities and continuing education programs should introduce practical *AI* evaluation modules, using frameworks like *CHAIR* instead of complex mathematics.[3] For governance, every *AI* tool requires a designated manager, performance tracking, regular audits, and inclusive approval committees.[4] Additionally, evidence-based practitioners should help design clinical trials and advocate for independent testing.[37]

In Türkiye, researchers must collaborate to build diverse datasets using privacy-preserving federated learning, allowing hospitals to train shared models without transferring patient images.[40] Furthermore, clinics must re-test imported *AI* tools locally before use, and developers must establish strict safety benchmarks for Turkish-language *AI* compliant with the Personal Data Protection Law (*KVKK*).[35]

Relevance to Clinical Practice

This review provides direct practical value for clinical workflows. The *CHAIR* checklist offers a structured, non-technical framework for evaluating commercial *AI* tools. Furthermore, it establishes realistic expectations: while current *AI* excels at narrow diagnostic tasks, published accuracy frequently degrades in new environments, making independent local validation essential before adoption. Third, responsibility remains with the dentist, who keeps the final decision, records disagreements and watches for drift and inequity once a tool is live. A practitioner who adopts these habits can take up useful tools earlier and more safely, defer adoption of tools with insufficient evidence, and take part in the local governance that decides which systems reach the chair.

Limitations

A few limitations should be acknowledged. As a narrative synthesis, the choice of studies followed no preregistered protocol and no formal appraisal of bias, so it is unavoidably a matter of judgement. Because the field advances so fast, the approval status and measured accuracy of any particular tool may well have moved on by the time these words are read. The coverage also leans towards image-based work, which is the most developed, and so probably underplays the rapid progress of language and foundation models. Finally, *CHAIR* is intended as a practical aid rather than a validated instrument and would itself benefit from testing in real clinical decisions. Its five domains were likewise derived from recurring failure points in the published literature rather than through a formal consensus method such as a *Delphi* process, a further reason to treat the checklist as a starting point for appraisal rather than a standard.

Conclusion

Artificial intelligence is no longer a distant prospect in dentistry but a present and uneven reality, most visible at the radiograph and the photograph and now reaching the conversation with the patient. The shift required of the profession is as much conceptual as technical. *AI is best understood as augmented intelligence*: narrow in scope and only as good as the data and labels behind it, and a new tool deserves the same scrutiny as any diagnostic test, namely a requirement for evidence gathered in comparable patients, for candid reporting of accuracy, for attention to bias and for a clear plan of oversight once it is in use. The *CHAIR* framework is meant to make that judgement structured and practical. Dentists who take part early, in choosing, testing and governing these systems rather than adopting them only after deployment, will help ensure that the tools reaching the chair fit their patients, their equipment and their way of working.

Acknowledgements

The author thanks to Associate Professor Fatma Sevinç Şengül, MD, for informal clinical feedback before preparation of the manuscript. The person named has seen and approved this acknowledgement. The author sincerely thanks the anonymous reviewers for their careful reading and constructive comments, which have substantially improved the quality and clarity of this manuscript.

Ethical Approval

No patients, human participants, identifiable personal data or animals were involved. Therefore, ethical approval was not required.

Conflicts of Interest

The author declares no conflicts of interest.

Sources of Funding

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

Authorship Contributions

Idea/Concept: D.Ş Design: D.Ş Control/Supervision: D.Ş Materials: D.Ş Data Collection and/or Processing: D.Ş Literature Review: D.Ş Analysis and/or Interpretation: D.Ş Writing the Article: D.Ş Critical Review: D.Ş

References

1. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25:44-56.
2. Schwendicke F, Samek W, Krois J. Artificial intelligence in dentistry: chances and challenges. *J Dent Res*. 2020;99(7):769-74.
3. Khanagar SB, Al-Ehaideb A, Maganur PC, et al. Developments, application, and performance of artificial intelligence in dentistry: a systematic review. *J Dent Sci*. 2021;16(1):508-22.
4. Mörch CM, Atsu S, Cai W, et al. Artificial intelligence and ethics in dentistry: a scoping review. *J Dent Res*. 2021;100(13):1452-60.
5. Moor M, Banerjee O, Abad ZSH, et al. Foundation models for generalist medical artificial intelligence. *Nature*. 2023;616(7956):259-65.
6. Cantu AG, Gehrung S, Krois J, et al. Detecting caries lesions of different radiographic extension on bitewings using deep learning. *J Dent*. 2020;100:103425.
7. Lee S, Oh SI, Jo J, et al. Deep learning for early dental caries detection in bitewing radiographs. *Sci Rep*. 2021;11:16807.
8. Bayrakdar IS, Orhan K, Akarsu S, et al. Deep-learning approach for caries detection and segmentation on dental bitewing radiographs. *Oral Radiol*. 2022;38(4):468-79.
9. Ammar N, Kühnisch J. Diagnostic performance of artificial intelligence-aided caries detection on bitewing radiographs: a systematic review and meta-analysis. *Jpn Dent Sci Rev*. 2024;60:128-36.
10. Krois J, Ekert T, Meinhold L, et al. Deep learning for the radiographic detection of periodontal bone loss. *Sci Rep*. 2019;9:8495.
11. Chang HJ, Lee SJ, Yong TH, et al. Deep learning hybrid method to automatically diagnose periodontal bone loss and stage periodontitis. *Sci Rep*. 2020;10:7531.
12. Khubrani YH, Thomas D, Slator PJ, White RD, Farnell DJJ. Detection of periodontal bone loss and periodontitis from 2D dental radiographs via machine learning and deep learning: systematic review employing APPRAISE-AI and meta-analysis. *Dentomaxillofac Radiol*. 2025;54(2):89-108.
13. Orhan K, Bayrakdar IS, Ezhov M, Kravtsov A, Özyürek T. Evaluation of artificial intelligence for detecting periapical pathosis on cone-beam computed tomography scans. *Int Endod J*. 2020;53(5):680-9.
14. Cui Z, Fang Y, Mei L, et al. A fully automatic AI system for tooth and alveolar bone segmentation from cone-beam CT images. *Nat Commun*. 2022;13:2096.
15. Tuzoff DV, Tuzova LN, Bornstein MM, et al. Tooth detection and numbering in panoramic radiographs using convolutional neural networks. *Dentomaxillofac Radiol*. 2019;48(4):20180051.
16. Vinayahalingam S, Xi T, Bergé S, et al. Automated detection of third molars and mandibular nerve by deep learning. *Sci Rep*. 2019;9:9007.
17. Fukuda M, Inamoto K, Shibata N, et al. Evaluation of an artificial intelligence system for detecting vertical root fracture on panoramic radiography. *Oral Radiol*. 2020;36(4):337-43.
18. Fu Q, Chen Y, Li Z, et al. A deep learning algorithm for detection of oral cavity squamous cell carcinoma from photographic images: a retrospective study. *EClinicalMedicine*. 2020;27:100558.
19. Warin K, Limprasert W, Suebnukarn S, Jinaporntham S, Jantana P. Performance of deep convolutional neural network for classification and detection of oral potentially malignant disorders in photographic images. *Int J Oral Maxillofac Surg*. 2022;51(5):699-704.
20. Hwang HW, Park JH, Moon JH, et al. Automated identification of cephalometric landmarks: part 2: might it be better than human? *Angle Orthod*. 2020;90(1):69-76.
21. Polizzi A, Leonardi R. Automatic cephalometric landmark identification with artificial intelligence: an umbrella review of systematic reviews. *J Dent*. 2024;146:105056.
22. Jung SK, Kim TW. New approach for the diagnosis of extractions with neural network machine learning. *Am J Orthod Dentofacial Orthop*. 2016;149(1):127-33.
23. Setzer FC, Li J, Khan AA. The use of artificial intelligence in endodontics. *J Dent Res*. 2024;103(9):853-62.

24. Sukegawa S, Yoshii K, Hara T, et al. Deep neural networks for dental implant system classification. *Biomolecules*. 2020;10(7):984.

25. Umer F, Batool I, Naved N. Innovation and application of large language models (LLMs) in dentistry: a scoping review. *BDJ Open*. 2024;10:90.

26. Tussie C, Starosta A. Comparing the dental knowledge of large language models. *Br Dent J*. 2024.

27. Huang H, Zheng O, Wang D, 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.

28. Aggarwal R, Sounderajah V, Martin G, et al. Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. *NPJ Digit Med*. 2021;4:65.

29. Nagendran M, Chen Y, Lovejoy CA, et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. *BMJ*. 2020;368:m689.

30. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med*. 2019;17:195.

31. 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):447-53.

32. Gichoya JW, Banerjee I, Bhimireddy AR, et al. AI recognition of patient race in medical imaging: a modelling study. *Lancet Digit Health*. 2022;4(6):e406-14.

33. Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. *Lancet Digit Health*. 2021;3(11):e745-50.

34. European Commission. Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Access date: June 20, 2026. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

35. Kişisel Verileri Koruma Kurumu. Personal Data Protection Law (Law No. 6698). Access date: June 20, 2026. https://www.kvkk.gov.tr/

36. U.S. Food and Drug Administration. Artificial intelligence-enabled medical devices. Access date: June 20, 2026. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-enabled-medical-devices

37. Liu X, Cruz Rivera S, Moher D, et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med*. 2020;26(9):1364-74.

38. Mongan J, Moy L, Kahn CE Jr. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): a guide for authors and reviewers. *Radiol Artif Intell*. 2020;2(2):e200029.

39. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. 2024;385:e078378.

40. Rischke R, Schneider L, Müller K, Samek W, Schwendicke F, Krois J. Federated learning in dentistry: chances and challenges. *J Dent Res*. 2022;101(11):1269-73.