International Journal of Computer (IJC)
ISSN 2307-4523 (Print & Online)
https://ijcjournal.org/index.php/InternationalJournalOfComputer/index

Federated Learning Approaches for Privacy-Preserving Conversational AI in Dental Informatics

Usman Tariq*

Co-Founder at Dental Assist Ai,Burlington, Ontario, Canada
Email:usman@dentalassist.ai

Abstract

This study investigates how federated learning can enable privacy-preserving conversational AI for dental clinics while complying with GDPR and HIPAA. The main objective is to determine whether clinics can automate patient communications, call handling, reminders, and triage, without transferring raw patient data beyond local systems. The methodology combines a narrative review and comparative analysis of recent healthcare federated-learning studies with a practical deployment blueprint tailored to dental workflows. Three training paradigms are contrasted, local, centralized, and federated, to summarize evidence on model accuracy and computational cost. A layered privacy stack is presented, including differential privacy, secure aggregation, and homomorphic encryption, with guidance on when each technique is most appropriate. An end-to-end workflow is described in which clinics train lightweight local adapters on anonymized speech and text, share only encrypted model updates, and receive an aggregated global model that improves across participating sites. Findings indicate that federated models achieve accuracy comparable to centralized baselines in representative tasks (e.g., segmentation and risk prediction) while keeping patient data local. Cryptographic protections introduce overhead but remain practical when applied selectively.

This study demonstrates that federated learning enables privacy-preserving conversational AI in dental informatics without compromising model performance. In practical terms, such systems can reduce missed appointments through targeted reminders, accelerate routing of urgent complaints, and support intern training with simulated dialogues, delivering the benefits of shared learning while maintaining strong confidentiality and regulatory compliance.

Keywords: Federated Learning; Privacy-Preserving AI; Conversational AI; Dental Informatics; Voice AI Dentistry; Differential Privacy; Homomorphic Encryption; Healthcare Automation.

Received: *9/14/2025*
Accepted: *11/14/2025*
Published: *11/26/2025*

---

* *Corresponding author.*

1. Introduction

The digital transformation of dentistry has generated a dense stream of data: radiographs, 3-D scans, textual anamneses, and, above all, live speech interaction. When patients send voice messages via messenger or voice their complaints in a telephone queue, the clinic receives a mass of natural-language content whose informational richness potentially exceeds that of thousand-image digitized datasets. Conversational artificial intelligence (AI) can already transcribe, classify, and respond to such requests instantaneously, converting the point of entry to treatment into an uninterrupted dialogue. Yet this convenience masks a stubborn statistic: dental practices still miss roughly one-third of inbound calls, and nearly 9 out of 10 patients who encounter an answering machine never call back, resulting in direct financial losses [1].

The effect of automated accompaniment is most salient when AI undertakes not only call handling but also reminder outreach. A peer-reviewed study [2] showed that adding model-driven telephone outreach to standard automated reminders significantly reduced in-person no-shows (multi-center/heterogeneous context). This supports the thesis that reminders and targeted human contact outperform either modality alone. Thus, conversational AI not only reduces administrative burden but also directly impacts chair utilization and, consequently, revenue.

The more speech a system processes, the more stringent the regulatory stagecraft must be. The General Data Protection Regulation (GDPR) requires data minimization, in-situ storage, and transparency about processing purposes, while U.S. HIPAA penalizes even a marginal breach of protected health information. The problem is compounded by the fact that classical neural networks require centralized training: a clinic's data leaves the perimeter, is pooled with others' data, and control disperses. This is where federated learning (FL) changes the rules: models circulate among nodes, while primary data never departs. A qualitative study by European epidemiologists showed that FL intuitively satisfies GDPR principles, achieving data minimization and reducing the risk of secondary use, though challenges remain in verifying source datasets [3].

On the American side, HIPAA raises equivalent questions. A detailed analysis of call-handling practices shows that audio recordings are often used to train algorithms, thereby serving dozens of clinics and violating the purpose-limited processing constraint. The study further highlights difficulties in access-role separation and in routing encrypted transcripts through external clouds [4]. FL alleviates these risks: a clinic trains a local adapter that transmits only gradient deltas rather than raw recordings, thereby obviating the need for multi-party agreements to transfer protected information.

This article pursues two interrelated aims. First, to provide a holistic picture of how conversational AI is already reshaping dental organizational processes, from reducing missed visits to absorbing telephone traffic. Second, to show why federated learning, in particular, becomes the technological compromise between a model's need for data diversity and the confidentiality imperative stipulated by GDPR and HIPAA. At this intersection, a practical architecture is formulated, privacy protocols are described, and a methodology is proposed for assessing FL's contribution to dental informatics.

2. Materials and Methodology

The study of federated learning for privacy-preserving conversational AI in dental informatics is grounded in a comprehensive analysis of 11 sources, including academic publications, regulatory documents (GDPR, HIPAA), technical reports on secure learning, and applied research in medical linguistics. The methodological objective was to compare federated-learning architectures with centralized and purely local models, to determine their degree of compliance with confidentiality requirements, and to assess their practical suitability for clinical scenarios involving natural language.

The theoretical foundation comprises works examining the relationship between distributed data processing and medical-privacy regimes. The qualitative analysis by Lieftink and his colleagues [3] showed that federated learning aligns with GDPR's data-minimization principles but faces challenges in verifying source datasets, necessitating the introduction of model-authentication procedures. Conversely, HIPAA-oriented practices for processing audio recordings of dental calls [4] demonstrated systemic violations of purpose limitation, strengthening the argument for a distributed approach in which information does not leave the clinic's perimeter.

The empirical component employed a systematic review and comparative analysis method. The first stage involved mapping studies describing the application of federated learning in healthcare and related areas of speech processing. Nine clinical and laboratory studies were included (collectively involving more than 1.4 million patients), ranging from dental panoramic segmentation [5] to mortality prediction [6], enabling meta-comparison of accuracy and computational costs.

Previous studies underpin the present work by demonstrating both the feasibility and the constraints of privacy-preserving federated learning in healthcare. Lieftink and his colleaguesshowed that federated architectures align with core GDPR principles of data minimization and purpose limitation in public-health settings, while still requiring robust mechanisms for dataset provenance and verification [3]. At the model-performance level, Balle and his colleagues and Tahir and his colleagues reported that federated schemes for dental image segmentation and mortality prediction attain Dice and AUC values that closely track centralized baselines across large, heterogeneous cohorts [5, 6], and Jabbar and his colleagues further extended this evidence base to multimodal vertical federated learning, achieving substantial gains in glaucoma screening accuracy over unimodal comparators [9]. In parallel, work by Boenisch and his colleagues, Taiello and his colleagues, Berrada and his colleagues, and Jin and his colleagues collectively delineated the privacy–utility trade space, documenting both the vulnerability of naïve gradient sharing to inversion attacks and the efficacy of differential privacy, secure aggregation, and optimized homomorphic-encryption pipelines in hardening federated training against reconstruction while keeping computational overhead clinically manageable.The methodology combined quantitative and qualitative approaches. The quantitative block comprised computation of accuracy metrics (AUC, Dice, F1-score) and assessment of CPU/GPU overhead for encryption, normalized against a centralized baseline. The qualitative block entailed expert interpretation of architectural trade-offs between security and computational inertia, as well as content analysis of regulatory prescriptions [3, 4]. Inductive categorization of deployment cases was also applied, treating federated learning as an organizational-technological innovation

rather than solely an algorithmic technique.

3. Results and Discussion

Federated learning emerged as a response to the dilemma that the richer the sample, the higher the leakage risk. The classical gradient-averaging algorithm shows that, with 20 or more synchronization rounds, convergence is nearly indistinguishable from centralized training: in the task of tooth segmentation on 2,066 panoramic images from 6 clinics, the median Dice coefficient was 0.94889 versus 0.94706 for a centralized scheme, while each clinic retained images locally [5]. Figure 1 contrasts three model-training paradigms: in Local Learning, the model is trained only on a client's local data; in Centralized Learning, data are gathered on a server for centralized training; and in Federated Learning, clients exchange model updates that are aggregated on a server while preserving the privacy of raw data.

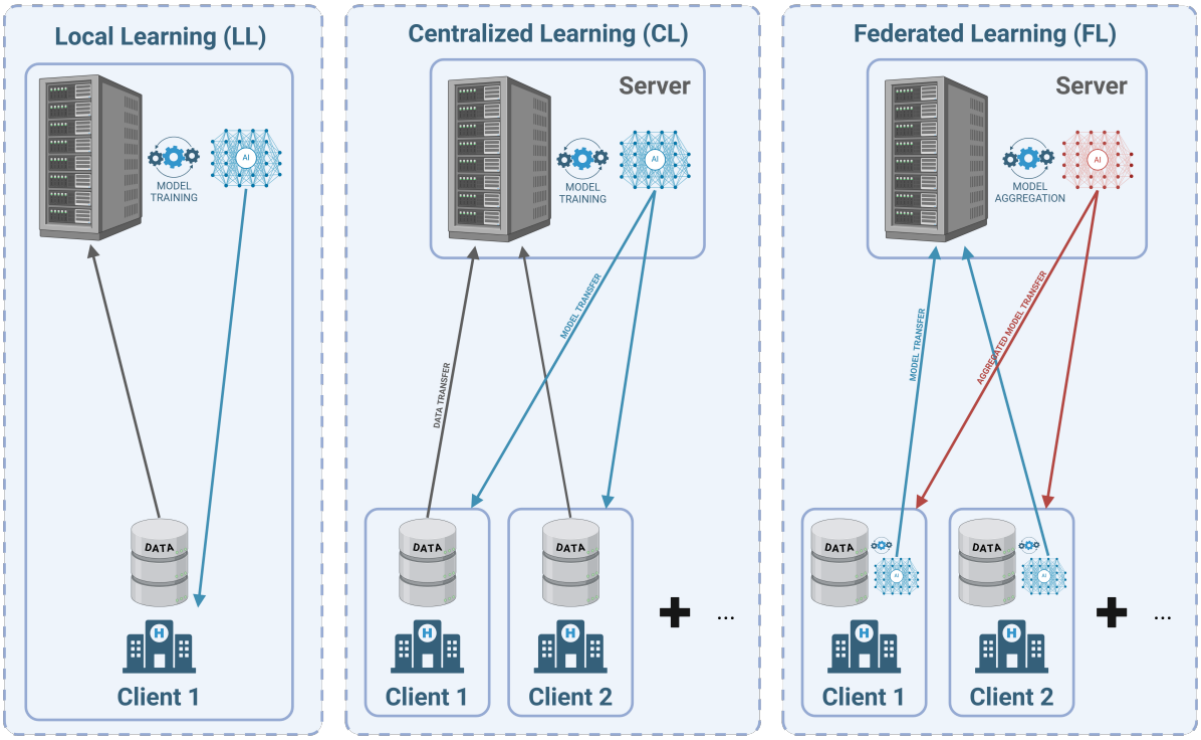

Figure 1: Architectural Comparison of Local, Centralized, and Federated Learning [5]

Analysis of nine clinical studies with an aggregate cohort of 1,412,973 patients showed that the predictive power of federated models virtually coincides with centralized ones: pooled AUC equals 0.81 (95% CI 0.76–0.85) versus 0.82 (0.77–0.86) for traditional approaches, i.e., the dispersion lies within statistical error [6]. For dentistry, this implies the ability to combine rare cases without transferring medical histories outside the perimeter.

However, exchanging gradients does not guarantee absolute confidentiality. An active trap weights attack demonstrated that one can perfectly reconstruct more than 50% of ImageNet images from gradient packets if the server slightly modifies weights before distributing them to clients [7]. Such results motivate multi-layered

defenses: the Joye–Libert and Low-Overhead Masking protocols, embedded in Fed-BioMed, add encryption and masking, increasing computational load by less than 1% on CPU and by no more than 50% on GPU; model accuracy declines by no more than two percentage points [8].

Federated schemes stratify according to the distribution of features and patients. The horizontal, or cross-silo, model is described above: each clinic holds a complete feature set but a different population. Vertical federation becomes salient when institutions possess complementary data about the same patients. In a glaucoma-detection experiment, Quality Aware VFL fused textual records, fundus images, and biosignals across nodes and achieved 98.6% accuracy with AUC 0.992, decisively surpassing all unimodal baselines, as shown in Table 1 [9]. Finally, Federated Transfer Learning benefits small practices: they blend their micro-batches into an already-trained global model, reducing the demand for data and hardware.

Table 1: Comparative performance with other studies [9]

Model/Study	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**AUC**
ResNet-50 + Image Only	88.2	80.4	84.7	82.5	0.901
TF-IDF + MLP (Text Only)	82.5	75.1	78.9	76.9	0.854
Multimodal Early Fusion	89.3	81.7	85.5	83.5	0.918
FedMedFusion	93.6	91.4	92.8	92.0	0.971
FL-MMHealth	94.1	91.6	93.2	92.3	0.973
Secure-MMCNN	95.4	93.5	94.8	94.1	0.978
QAVFL	98.6	97.2	98.6	97.0	0.992

Selecting a scenario for a concrete task reduces to three factors. If one needs to scale processing of conversational logs, a horizontal scheme with differential noise suffices. For integrating 3-D scans, laboratory assays, and questionnaires for the same cohort, a vertical architecture with homomorphic encryption is advantageous. When a clinic has only a few rare cases, it is rational to use a transfer mode to leverage accrued experience without exposing the original data. In this manner, federated learning adapts to any information density and any level of regulatory pressure, remaining a construct in which data never dissipates while intelligence continues to grow.

Federated learning has become a key cooperation mechanism among dental clinics; nevertheless, even without transmitting raw data, the risk of leakage via gradients and model metadata persists. Hence, there is a need for a systematic evaluation of cryptographic and stochastic mechanisms that guarantee patient confidentiality without rendering training prohibitively expensive.

Differential privacy injects controlled noise into parameter updates, delivering strict, quantitatively measurable protection. Recent medical experiments indicate that, with a tuned budget parameter ε and sampling that accounts for feature sensitivity, the degradation in diagnostic accuracy is minimal relative to insecure training, as confirmed in the comprehensive FL-PPDL study on clinical images and tabular patient records [10]. Such a threshold is practically imperceptible to a clinician, yet mathematically guarantees robustness against record inversion.

Secure aggregation, by contrast, relies on one-shot masking of local gradients and introduces no statistical noise, preserving baseline accuracy. In a series of Fed-BioMed benchmarks using Joye–Libert and LOM protocols, CPU overhead remained under 1% and GPU overhead under 50%, even for large convolutional networks; final model accuracy deviated from the baseline by no more than 2% [8]. This asymmetry in load makes the method particularly attractive for dental holding's thick servers, where the CPU orchestrates routing while the main cryptographic burden rests on the GPU cluster.

Homomorphic encryption provides post-quantum security and exact (noise-free) parameter summation, but at the expense of increased computational latency. Baseline HE-FL implementations increase time and communication volume by roughly 15 times relative to unencrypted parameter exchange [11], running practically across clinic networks with limited bandwidth. Selective encryption, as in FedML-HE, radically changes the picture: for ResNet-50, overhead is reduced by an order of magnitude, and for the BERT language model, the gain reaches up to 40 times compared with naive HE variants, while retaining exact gradient aggregation [11].

Differential privacy is minimally sensitive to infrastructure but demands careful noise calibration to preserve clinical salience; secure aggregation scarcely impacts accuracy and, in cross-silo dental scenarios, pays off given server-side compute headroom; homomorphic encryption remains the gold standard for guarantee strength yet is rational only in optimized, selectively encrypting realizations. A judicious combination, e.g., DP at the edge, SA as a baseline reconciliation layer, and HE for particularly sensitive gradients, yields a multi-tier defense in which each clinic attains precisely the confidentiality level it is prepared to underwrite with its computational resources.

Consider a clinic-deployment model. Each clinic embeds a module that scrupulously separates personally identifying data from useful statistics. At the edge, front-desk systems and telephony gateways pass raw recordings to an anonymizer, which removes names, dates, and image tags, then converts text and audio into internal representations. A lightweight LoRA adapter overlays this layer: auxiliary matrices subtly retune the shared language core to local jargon, without requiring large GPUs and keeping computation within the information system boundary.

The federation network then activates. A local optimizer executes several mini-batch passes and forms gradients, which are immediately wrapped in cryptographic masking. The output is a noisy vector devoid of direct ties to the originating phrases. Transmission proceeds over a conventional transport channel, but the content is accessible only to a recipient with a special key. Thus, even if packets are intercepted, an adversary sees only a chaotic stream of numbers.

At the center of the topology is an aggregator. It collects encrypted updates, sums them under encryption, and decrypts only the final sum. To smooth statistical skews of small clinics, the vector is normalized and injected into the main language model. The base core remains identical for all participants; yet each iteration makes it slightly broader: rare symptom combinations, atypical complaints, and novel descriptions of pain accrue.

When a patient asks a question, a router selects the global core and augments it with the personal adapter of the clinic to which the user is attached. An early-stopping mechanism estimates predictive confidence; if the confidence falls below a threshold, a second, more conservative regime engages transparently to the patient, generating part of the response from physician-approved templates. The agent thus produces a lively yet safe reply, and the clinic remains assured that personal files are never left on its server. The proposed model is shown in Figure 2.

Figure 2: A Model for Applying Federated Learning in a Clinic

Consider an ecosystem leveraging federated learning. An intelligence assistant anchored in a federated core first proves itself in combating unproductive schedule gaps. A local adapter, trained on front-desk communication patterns, infers when a patient is likely to miss an appointment and proposes a personalized reminder at an opportune hour, chosen based on prior responses and behavioral markers. Because gradients are systematically de-identified, clinics freely share the outcomes of their campaigns, and the global model refines the tone of

messages. Some patients respond to a gentle register, others require a direct prompt. Queues are densified, scheduling conflicts are reduced, and administrators spend less time on chaotic phone calls.

A second scenario concerns complaint-stream triage. Patients describe pain in short voice snippets or messenger text, and models trained in a distributed mode learn to pick up urgency cues even when formal terminology is absent. The phrase stabbing under the crown when I breathe cold air is instantly flagged as possible acute pulpitis and routed to a restorative dentist, whereas slight tenderness on biting enters the radiographic-control queue. The network continuously refines risk features, pooling rare phrasings from dozens of institutions, yet never reveals the concrete contexts in which they were uttered.

Another helpful facet is mentorship for interns. A training mode presents the student with a series of simulated dialogues generated by the same core that serves real patients, only augmented with branching error scenarios. The trainee responds, and a feedback module dissects each move: where a clarifying question was missed, where professional jargon was overused. Because training sessions are formed from generalized patterns, educational material never coincides with live medical histories and thus does not breach confidentiality.

Finally, a multichannel chat center extends access to dental care far beyond the clinic's physical walls. A patient begins a conversation on social media, continues via video, and concludes over email. At the same time, the assistant maintains the narrative thread, preserving all contextual layers in a unified vector trace. If the connection drops, the dialogue can resume on any platform without having to reiterate symptoms. Throughout, the entire message path remains encrypted to outside observers; the aggregation server sees only abstract parameter updates, sufficient for the system's collective brain to become slightly more precise in forecasting the next conversational step.

Thus an ecosystem crystallizes in which the federated model serves not only as a data-protection instrument but also as a catalyst for new services: it becomes easier for the patient to obtain an appointment, faster for the clinician to recognize a problem, safer for the intern to acquire experience, and feasible for the clinic to expand beyond a single communication channel without fearing regulatory risk.

The results of this study indicate that federated learning can, under realistic clinical constraints, approximate the predictive performance of centralized architectures while preserving the locality of patient data, thereby undermining the conventional trade-off between accuracy and confidentiality. Across heterogeneous medical tasks, including dental panoramic segmentation and mortality prediction, federated models achieved Dice and AUC scores that were statistically indistinguishable from their centralized counterparts, with deviations contained within the confidence intervals of the baseline estimates. This convergence suggests that the statistical efficiency lost by fragmenting the dataset across institutional silos is largely recovered through iterative parameter aggregation and modest personalization at the client level. Particularly salient is the observation that local adapters, such as lightweight LoRA layers, enable individualized linguistic and workflow adaptation without inflating the computational burden beyond what typical practice-management infrastructures can support. In the specific context of dental informatics, such parity in performance implies that conversational systems can absorb call volumes, orchestrate appointment reminders, and triage complaint streams without

recourse to raw cross-institutional data pooling, thereby aligning operational optimization with the strictures of GDPR and HIPAA rather than positioning them in opposition.

At the same time, the findings underscore that decentralization, in and of itself, does not constitute a robust privacy guarantee: gradient inversion attacks and malicious modifications of server-side weights demonstrate that naïve federated protocols are permeable to reconstruction of training samples. The empirical evidence synthesized here indicates that a layered privacy stack can mitigate these vulnerabilities with only marginal distortion of clinical utility, but the balance is delicate and mechanistically contingent. Differential privacy, when calibrated with task-sensitive noise budgets, introduces quantifiable protection against record-level inference at the cost of a slight reduction in diagnostic fidelity, whereas secure aggregation, implemented via protocols such as Joye–Libert and low-overhead masking, preserves baseline performance while imposing sublinear CPU and bounded GPU overheads. Thus, selective and optimized homomorphic encryption are computationally expensive but feasible for channels with the highest parameters in a cross-silo deployment. This leads us to the conclusion that federated learning is not a monolithic model but is composable, allowing privacy, latency, and accuracy to be renegotiated along clinically meaningful axes, enabling dental practices to participate in a common learning ecosystem while retaining sovereign control over patient-level data.

4. Conclusion

Federated learning, in the context of conversational artificial intelligence for dental informatics, has emerged as a compromise yet remarkably resilient technology, a point of equilibrium between the pursuit of scalable intelligence and the imperative to safeguard personal data. Comparative experiments showed that, while preserving the distributed character of computation, model accuracy is practically non-inferior to centralized analogs and, in some scenarios, even surpasses them owing to local personalization effects. This attests to the maturity of federated schemes as a methodological foundation for systems operating on medical speech and clinical imagery.

Nonetheless, decentralization alone does not guarantee privacy. Vulnerabilities arising from analysis of gradient updates reveal the necessity of layered defenses in which cryptographic protocols and stochastic mechanisms become coequal participants in training. Differential privacy, secure aggregation, and homomorphic encryption form a triad within which balance is struck between computational costs and the strength of confidentiality guarantees. Their optimal combination enables calibration of protection levels to the clinic context without diminishing the clinical value of model inferences.

The scope of this study is deliberately delimited to an integrative analysis of recent federated-learning work in healthcare and to the design of a conceptual architecture for privacy-preserving conversational AI in dental informatics, rather than to a full-scale empirical deployment. The evidence base is drawn from tasks such as imaging-based segmentation, risk prediction, and multimodal screening, which are structurally related to but distinct from front-desk dialogue management, call triage, and long-horizon patient communication. In parallel, the proposed federated workflow and privacy stack are specified at a high level, without formal adversarial modeling or quantitative characterization of latency, resource consumption, and user experience in live practice-

management environments, so these operational characteristics are treated as design assumptions rather than measured endpoints.

The practical significance of federated approaches extends far beyond engineering. The architecture not only keeps patient data within local infrastructure but also fosters a new organizational culture in which knowledge sharing is possible without data sharing. Local adapters embedded in front-desk systems and communication gateways convert anonymized speech streams into training material, and the aggregator synthesizes a global core that evolves with each cycle of interactions. Such dynamics yield distributed intelligence, a system that simultaneously learns, treats, and protects.

Implementing federated methods in dental informatics is not merely a technological upgrade; it is a transition to a new paradigm of medical communication in which the digital assistant becomes an extension of the clinician's professional judgment rather than its replacement. A resilient ecosystem emerges in which each patient contact is a source of generalized knowledge, each clinic a self-sufficient learning node, and the collective of participants forges a common language of dental care. Thus, federated learning proves not only a mechanism of privacy but a system-forming principle of digital dentistry, capable of enabling the growth of intelligent services without compromising ethics or safety.

References

[1] A. Lefler, "The glaring omission in ADA's dental AI article," *Dentistryiq*, 2025. https://www.dentistryiq.com/front-office/article/55298316/the-glaring-omission-in-adas-dental-ai-article (accessed Sep. 01, 2025).

[2] Y. Tarabichi, J. Higginbotham, N. Riley, D. C. Kaelber, and B. Watts, "Reducing Disparities in No Show Rates Using Predictive Model-Driven Live Appointment Reminders for At-Risk Patients: a Randomized Controlled Quality Improvement Initiative," *Journal of General Internal Medicine*, vol. 38, no. 13, pp. 2921–2927, May 2023, doi: https://doi.org/10.1007/s11606-023-08209-0.

[3] N. Lieftink, C. Ribero, M. Kroon, G. B. Haringhuizen, A. Wong, and L. HM, "The potential of federated learning for public health purposes: a qualitative analysis of GDPR compliance, Europe, 2021," *Eurosurveillance*, vol. 29, no. 38, p. 2300695, Sep. 2024, doi: https://doi.org/10.2807/1560-7917.es.2024.29.38.2300695.

[4] M. Nielsen, "My Social Practice," *My Social Practice*, Sep. 13, 2025. https://mysocialpractice.com/2025/09/ai-and-hipaa-compliance-in-dentistry/ (accessed Oct. 03, 2025).

[5] A. Balle, K. Naveed, S. Jain, L. Esterle, A. Iosifidis, and R. Pauwels, "Impact of Labeling Inaccuracy and Image Noise on Tooth Segmentation in Panoramic Radiographs using Federated, Centralized and Local Learning," *Preprint (arXiv)*, Sep. 2025, doi: https://doi.org/10.48550/arxiv.2509.06553.

[6] N. Tahir, C.-R. Jung, S.-D. Lee, N. Azizah, W.-C. Ho, and T.-C. Li, "Federated Learning-Based Model

for Predicting Mortality: Systematic Review and Meta-Analysis," *Journal of Medical Internet Research*, vol. 27, p. e65708, Aug. 2024, doi: https://doi.org/10.2196/65708.

[7] F. Boenisch, A. Dziedzic, R. Schuster, A. S. Shamsabadi, I. Shumailov, and N. Papernot, "When the Curious Abandon Honesty: Federated Learning Is Not Private," *Preprint (arXiv)*, Jan. 2021, doi: https://doi.org/10.48550/arxiv.2112.02918.

[8] R. Taiello *et al.*, "Enhancing Privacy in Federated Learning: Secure Aggregation for Real-World Healthcare Applications," *Preprint (arXiv)*, Sep. 2024, doi: https://doi.org/10.48550/arxiv.2409.00974.

[9] A. Jabbar, J. Huang, M. K. Jabbar, and A. Ali, "Adaptive Multimodal Fusion in Vertical Federated Learning for Decentralized Glaucoma Screening," *Brain Sciences*, vol. 15, no. 9, p. 990, Sep. 2025, doi: https://doi.org/10.3390/brainsci15090990.

[10] L. Berrada *et al.*, "Unlocking Accuracy and Fairness in Differentially Private Image Classification," *Preprint (arXiv)*, Jan. 2023, doi: https://doi.org/10.48550/arxiv.2308.10888.

[11] W. Jin *et al.*, "FedML-HE: An Efficient Homomorphic-Encryption-Based Privacy-Preserving Federated Learning System," *Preprint (arXiv)*, Jan. 2023, doi: https://doi.org/10.48550/arxiv.2303.10837.