multimedia

MDPI

Review

Privacy-Preserving Techniques for the Collection, Analysis, and Dissemination of Multimedia Data

Hassan Mistareehi [1,*], Md Sultan Al Nahian [2] and Zheng Zhang [1]

[1] Department of Computer Science & Information Systems, Murray State University, Murray, KY 42071, USA; zzhang22@murraystate.edu
[2] Department of Computer Science and Mathematics, Penn State Harrisburg, Middletown, PA 17057, USA; mkn5526@psu.edu
* Correspondence: hmistareehi@murraystate.edu

Abstract

The ubiquitous generation of multimedia data—images, video, and audio—by smartphones, surveillance networks, and wearable sensors has created unprecedented analytical opportunities while amplifying privacy risks. Unlike structured tabular data, identity in multimedia is redundantly embedded in pixels, waveforms, and metadata, rendering traditional de-identification methods brittle. Furthermore, modern deep learning models can re-identify individuals even after conventional redaction. This paper organizes privacy-preserving multimedia techniques across the data lifecycle: collection, analysis, and dissemination. We examine the principal technical families, including visual/acoustic obfuscation, differential privacy, federated and split learning, homomorphic encryption, secure multi-party computation, generative adversarial approaches for synthetic data, and blockchain-based access control. We synthesize the state of the art regarding the privacy–utility trade-off, summarize representative attacks (re-identification, membership inference, model inversion, and gradient leakage), and present a comparative taxonomy across modalities, lifecycle stages, and computational costs. We further offer a cross-cutting discussion identifying four themes—the insufficiency of any single mechanism in isolation, an under-addressed adaptive-adversary problem, an unresolved gap between formal privacy guarantees and perceptual/task utility, and uneven coverage across modalities—and conclude by identifying open challenges, including scalable encrypted video processing, standardized privacy–utility metrics, regulatory harmonization, and robustness against adaptive adversaries, that define the future research agenda.

Keywords: privacy-preserving computation; differential privacy; federated learning; homomorphic encryption; face de-identification; voice anonymization; secure multi-party computation; GDPR; multimedia security; synthetic data

Check for updates

Academic Editor: Goorak Kwon

Received: 24 July 2026
Revised: 18 September 2026
Accepted: 20 October 2026
Published: 26 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Multimedia data now constitutes the majority of global internet traffic, driven by smartphones, dashcams, body-worn cameras, smart speakers, and social media platforms [1]. This wealth of visual and acoustic data has catalyzed significant advances in computer vision, speech recognition, and multimodal artificial intelligence. However, these archives represent some of the most privacy-sensitive assets held by organizations today, and the pace at which they are generated is accelerating: consumer devices now default to continuous sensing (always-on cameras, ambient microphones, and wearable biosensors), which

means that privacy-preserving engineering can no longer be treated as an optional add-on bolted onto a finished analytics pipeline. It must be designed into the pipeline from the point of capture onward.

Unlike tabular data, where identity is tied to discrete, enumerable attributes (e.g., names, ID numbers), multimedia encodes identity implicitly and redundantly—through facial geometry, gait, voiceprints, license plates, background landmarks, and even reflections in eyeglasses or windows [2]. Moreover, embedded EXIF metadata, which often contains GPS coordinates and timestamps, adds an additional, frequently overlooked leakage channel [3]. Critically, the same deep learning advances that enable analytics also erode naive privacy protections: super-resolution networks can reverse pixelation, and large-scale face and speaker recognition systems can re-identify individuals in supposedly anonymized footage [4,5]. This creates a moving target for privacy engineering—a defense that is adequate against today's recognition models may be defeated by tomorrow's, a theme we return to throughout this review and treat explicitly as an open challenge in Section 11.

This risk is reflected in an increasingly stringent regulatory landscape. Frameworks such as the European Union's General Data Protection Regulation (GDPR), the U.S. Health Insurance Portability and Accountability Act (HIPAA), the California Consumer Privacy Act (CCPA), and China's Personal Information Protection Law (PIPL) explicitly classify biometric and health-related imagery as sensitive data requiring heightened protection [6]. Consequently, organizations—hospitals, smart cities, autonomous vehicle developers, and social media companies—must balance analytical utility with strict privacy obligations, often under threat of substantial financial and reputational penalties for non-compliance.

A useful organizational lens is the data lifecycle: (i) Collection, when raw signals are captured; (ii) Analysis, when data is processed and models are trained or queried; and (iii) Dissemination, when results, models, or raw data are shared with third parties. Each stage exposes a different attack surface and thus favors distinct defensive techniques: data minimization and on-device obfuscation dominate collection; differential privacy, federated learning, and secure computation dominate analysis; and anonymized or synthetic publishing, together with cryptographically governed access, characterizes dissemination. Figure 1 summarizes this lifecycle view and previews the technique families examined in Sections 3–5.

This review makes four contributions. First, it provides a structured account of privacy-preserving techniques organized explicitly by lifecycle stage rather than by mathematical mechanism alone, which we believe better reflects how practitioners actually design multimedia systems. Second, it synthesizes the attack literature—re-identification, membership inference, gradient and model inversion, and metadata/location inference—as the direct motivation for each defensive family, rather than treating attacks and defenses as separate bodies of work. Third, it presents a comparative taxonomy and accompanying visual analyses relating techniques to modalities, privacy–utility trade-offs, and computational cost, which we have not seen presented together elsewhere in this specific form. Fourth, it outlines seven concrete open research challenges, several of which (metric standardization, adaptive-adversary robustness, and privacy–fairness interaction) we argue are currently more consequential bottlenecks than the invention of further point techniques.

We substantiate this novelty claim rather than only asserting it. We identified three recent surveys whose scope overlaps with this review and compared them directly (Table 1): a visual-content privacy survey covering image and video but not audio, cryptographic computation techniques, or lifecycle organization; an image-only privacy survey with a narrower modality and technique scope; and a data-lifecycle survey that shares our organizing principle but is not multimedia-specific and does not address modality-specific

identity encoding or attacks [7–9]. Each of these three surveys covers at most one or two of the three dimensions—cross-modality coverage, lifecycle organization, and attack-to-defense synthesis—that this review combines; we did not identify a survey in our search that combines all three. We do not claim this comparison is exhaustive of the survey literature, only that it substantiates the specific novelty claims made above against the most directly overlapping work we found.

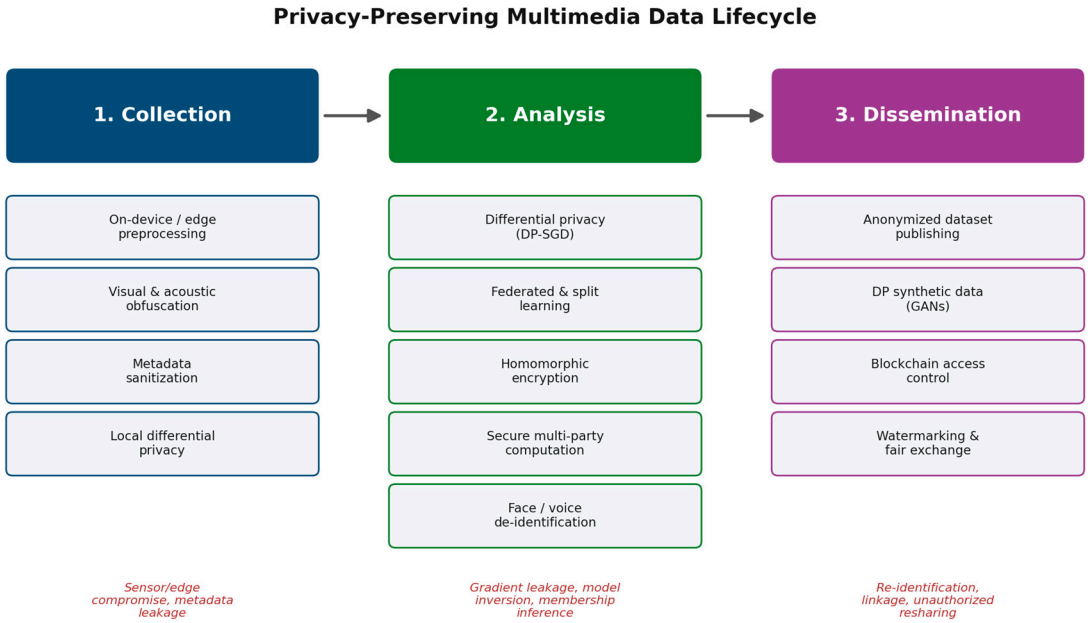


Figure 1. The privacy-preserving multimedia data lifecycle, with representative techniques and attack surfaces at each stage.

Table 1. Comparison of this review against recent, directly overlapping surveys.

Survey	Modalities Covered	Organizing Framework	Coverage Relative to This Review
Zhao et al., 2025 [7] (Visual Content Privacy Protection: A Survey, ACM Computing Surveys)	Image, video	Adversary type (human-vision vs. computer-vision) and protection mechanism (obfuscation, synthesis, encryption)	No audio; no differential privacy, federated/split learning, homomorphic encryption, secure multi-party computation, or blockchain-based access control; not organized by data lifecycle stage; no treatment of formal privacy guarantees (ε-DP, k-anonymity bounds).
Wen et al., 2026 [8] (Image Privacy Protection: A Survey, arXiv:2412.15228)	Image only	Encryption-based vs. non-encryption-based (visual) protection	No video or audio; no cryptographic computation techniques beyond image encryption (no HE, SMC, or FL); not organized by data lifecycle stage; attack coverage limited to image-specific re-identification.
Madhusudhanan and Jose, 2025 [9] (Privacy Preservation Techniques through Data Lifecycle, Computers and Security)	General/structured and unstructured data; not multimedia-specific	Data lifecycle (collection, storage, processing, sharing/transmission, retention/deletion, access/control)	Lifecycle-organized, like this review, but not multimedia-specific: does not address modality-specific redundant identity encoding in pixels or waveforms, multimedia re-identification attacks (e.g., gait or voice), or visual/acoustic obfuscation and face/voice de-identification techniques.
This review	Image, video, audio, and metadata jointly	Data lifecycle (collection, analysis, dissemination) combined with an explicit attack-to-defense mapping	Combines lifecycle organization, cross-modality coverage, and attack-to-defense synthesis in a single framework; no other survey we identified combines more than one of these three dimensions.

The remainder of the paper is organized as follows. Section 2 establishes the threat model, adversary taxonomy, regulatory background, and the review methodology used to identify and select the literature discussed. Sections 3–5 cover the three lifecycle stages—collection,

analysis, and dissemination—in turn. Section 6 reviews the privacy attacks that motivate these defenses in more technical depth. Section 7 provides the comparative taxonomy. Section 8 offers a cross-cutting discussion that synthesizes recurring themes and tensions across the preceding sections. Section 9 discusses five application domains in which these techniques are deployed jointly. Section 10 outlines open challenges and a future research agenda. Section 11 concludes the paper.

2. Background: Threat Model, Regulatory Context, and Review Methodology

Multimedia privacy differs from tabular-data privacy in three key ways. First, identity cues are high-dimensional and redundant: a single video clip reveals facial identity, age, gender, location, and socioeconomic markers simultaneously; removing one cue (e.g., blurring the face) leaves others—gait, clothing, voice, background objects—intact [2,10]. Second, multimedia is compressible by representation learning: deep neural networks extract compact embeddings that retain re-identifying power even when the original signal has been substantially degraded or ostensibly anonymized [4]. Third, multimedia pipelines are typically multi-stage and multi-party (e.g., edge capture, cloud storage, third-party analytics), requiring consistent privacy guarantees across trust boundaries that individually scoped techniques were not designed to span. Together, these properties explain why point solutions that work well for structured records—for instance, simple attribute suppression or k-anonymity over a handful of quasi-identifiers—transfer poorly to images, video, and audio without substantial modification.

Consistent with the broader privacy and security literature, this review distinguishes four adversary models, each of which motivates a different defensive posture and recurs throughout Sections 3–6. Honest-but-curious service providers execute the agreed computation correctly (e.g., they will genuinely train the model or run the query the data owner requested) but will attempt to learn additional information from intermediate values, model parameters, or outputs; this model motivates encryption-based techniques (homomorphic encryption, secure multi-party computation) and federation, where the provider never receives raw data at all. Linkage adversaries attempt to re-identify individuals in ostensibly anonymized content by combining it with auxiliary information—a second dataset, background knowledge, or a more capable recognition model than the one the defender evaluated against; this is the adversary model that k-anonymity-based anonymization (Section 4) is designed to address. As Section 5 discusses, however, k-anonymity bounds re-identification risk only with respect to the reference set and recognition capability assumed at design time; it is not a worst-case guarantee against an adversary with stronger auxiliary information or a more capable recognition model than the one originally evaluated against, and visual k-anonymity in particular has been shown to be defeated by exactly this kind of adversary [11]. Model-inversion and membership-inference adversaries query or otherwise access a trained model and attempt to infer properties of, or reconstruct, its training data; foundational work by Fredrikson et al. showed that confidence scores returned by a classifier can be exploited to reconstruct recognizable training-set face images, and follow-up work in federated settings showed that shared gradients themselves are frequently sufficient to reconstruct near-pixel-perfect training images [12,13]. This adversary class is the direct motivation for differential privacy. Finally, eavesdroppers intercept data in transit between parties and are addressed by conventional and homomorphic encryption. Figure 2 summarizes how each adversary maps onto a primary attack and the defense family it motivates; this mapping recurs as an organizing thread across Sections 3–6.

It is worth emphasizing that these adversary models are not mutually exclusive in a deployed system: a cloud analytics provider may be simultaneously honest-but-curious

with respect to the raw data it processes and a potential linkage adversary with respect to any anonymized outputs it later publishes. Robust system design therefore typically requires composing defenses drawn from more than one column of Figure 2, a point we return to in Section 8.

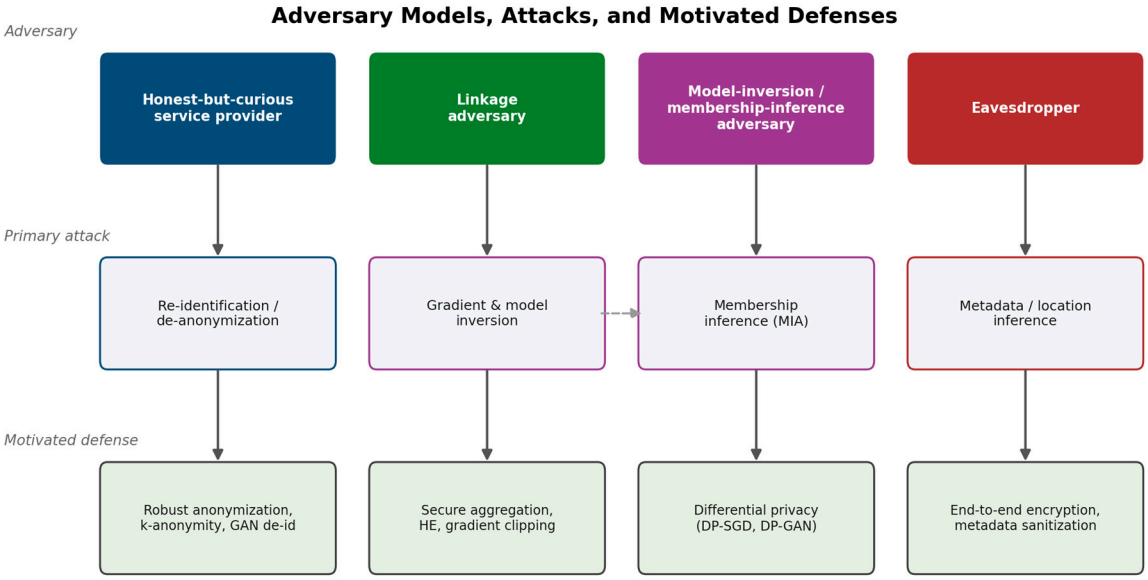


Figure 2. Adversary models, their primary attacks, and the defenses that address them; dashed lines mark closely related attack pairs.

GDPR Article 9 restricts the processing of biometric data used for the purpose of uniquely identifying a natural person, treating it as a special category of personal data; HIPAA governs identifiable health imagery in the United States; CCPA/CPRA provides consumers with rights over the collection and sale of personal information, including inferred profiles; and PIPL mandates consent and, in some cases, data localization within China [6]. These frameworks create concrete compliance pressure for technical mechanisms at each lifecycle stage: data minimization and purpose limitation at collection, privacy-preserving computation during analysis, and auditable, consent-respecting dissemination. Although a full legal analysis is outside the scope of this review, we note throughout Sections 3–5 where a given technical mechanism directly supports a specific regulatory obligation (e.g., federated learning supporting data-localization requirements or differential privacy supporting demonstrable, quantifiable de-identification). We emphasize that these are supportive relationships, not equivalences: a technique such as federated learning, differential privacy, or on-device processing can advance a specific regulatory objective, but satisfying that objective technically does not by itself constitute legal compliance with GDPR, HIPAA, CCPA/CPRA, or PIPL, each of which imposes additional procedural, documentation, and governance requirements beyond the technical mechanism itself. We therefore describe these connections throughout Sections 3–5 as mechanisms that support particular regulatory principles rather than as mechanisms that establish compliance.

Nearly every technique surveyed in this review trades analytical utility for privacy protection, and the shape of that trade-off—not merely its existence—is what differentiates competing approaches. For example, aggressive video blurring may protect identity but destroy the activity-recognition cues a downstream system needs; the severity of this trade-off depends on filter type, blur radius, and downstream task and is not a fixed property of "blurring" as a technique [14]. Formally, for a technique parameterized by a privacy budget or strength parameter (such as differential privacy's ε), utility is typically a monotonically non-increasing function of privacy strength, but the slope of that function

varies enormously across techniques and tasks—a fact obscured when privacy techniques are compared only by their formal guarantees rather than by measured downstream performance. This trade-off is a central axis of comparison in Section 7 and is visualized directly in Figure 3.

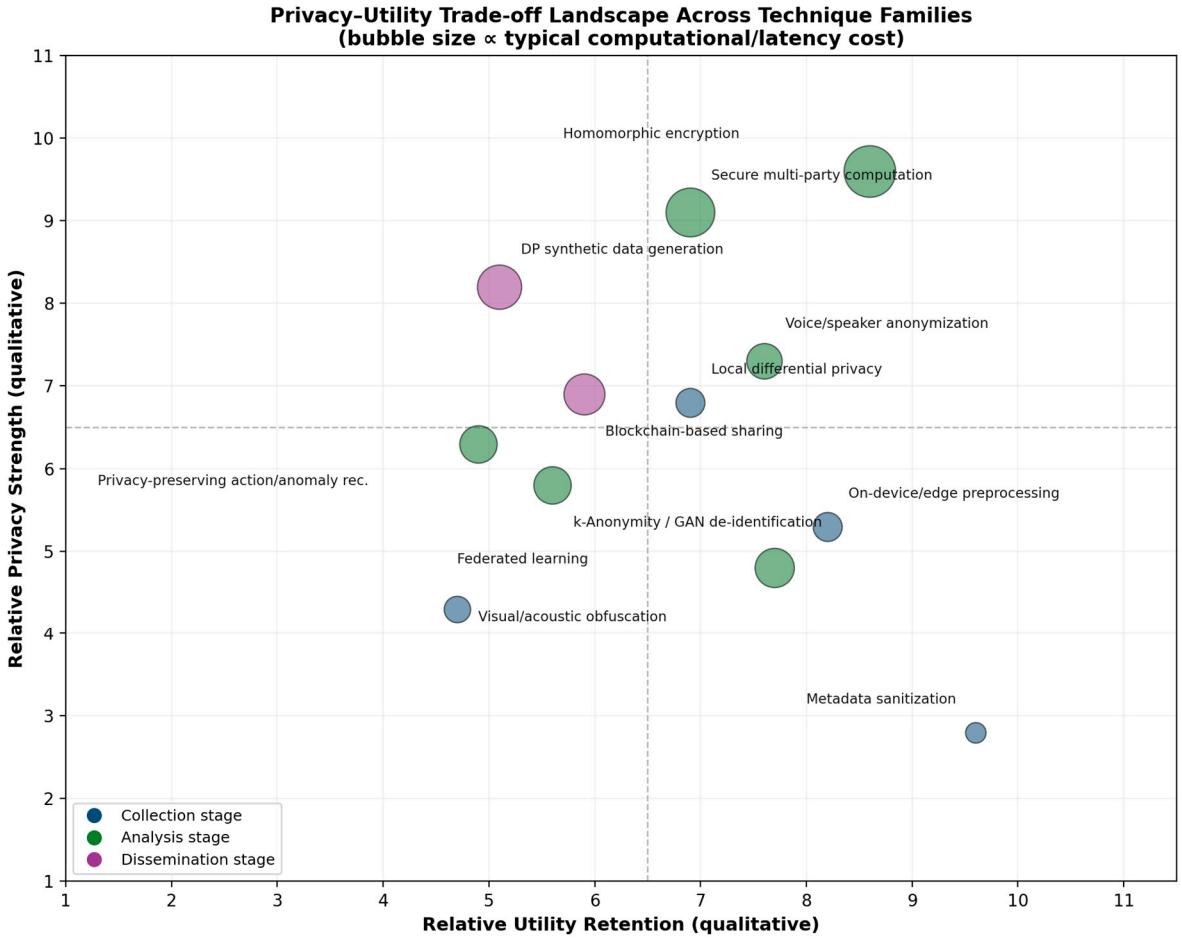


Figure 3. Privacy–utility trade-off landscape across the technique families reviewed in Sections 3–5.

Literature Search Strategy and Scope

Sources were identified through the databases, search-term families, and publication-year window below, applied iteratively across multiple search passes and supplemented by backward and forward citation chaining from key papers identified in each pass, rather than executed as a single fixed query.

- Databases queried: IEEE Xplore, the ACM Digital Library, SpringerLink, ScienceDirect, and arXiv, supplemented by targeted Google Scholar searches.
- Venues prioritized: peer-reviewed journal articles and top-tier conference proceedings (e.g., ACM CCS, IEEE S&P, NeurIPS, ICCV, CVPR, ICLR), together with relevant arXiv preprints, given the fast-moving nature of this field.
- Privacy-mechanism search terms: differential privacy, federated learning, homomorphic encryption, secure multi-party computation, de-identification, anonymization, synthetic data, and blockchain access control.
- Multimedia-modality search terms: image, video, face, speech, voice, biometric.
- Attack search terms: membership inference, model inversion, gradient leakage, and re-identification.
- Term combination: privacy-mechanism terms were paired with modality terms and, separately, with attack terms (e.g., “differential privacy” with “video”; “membership

inference" with "face"), since no single query spanning all three vocabularies at once returned usable results in these databases.

- Publication window: 2005–2026, capturing both foundational works (e.g., k-anonymity for faces, the theoretical foundations of differential privacy) and the most recent attack and defense literature available at the time of writing.

This strategy identifies a defensible, broad sample of the literature but falls short of a fully reproducible systematic search: we did not log the exact per-database Boolean query strings executed, the specific calendar dates on which each search was run, or a stage-by-stage count of records identified, screened, and included. Reconstructing these figures retroactively would risk misrepresenting the process as more formally controlled than it was, so we state this limitation directly rather than omit it or fabricate it.

We characterize this explicitly as a narrative rather than a systematic review: we did not apply a PRISMA-style protocol with a fixed final search date, a pre-registered inclusion/exclusion checklist, dual independent screening, or a recorded count of records identified, screened, and included at each stage, and we did not systematically track or report preprints that were considered but excluded. This choice reflects the breadth of the topic, which spans cryptography, machine learning, and multiple sensing modalities, and for which a single systematic search strategy across all sub-areas would be difficult to specify without either narrowing scope substantially or sacrificing the cross-modality synthesis that is the central contribution of this review. We flag this as a limitation rather than an oversight: readers should treat the coverage in Sections 3–6 as illustrative of the state of the field as identified through the databases, search terms, and time window described above, rather than as a formally exhaustive or independently reproducible census of the literature.

3. Privacy-Preserving Techniques at the Collection Stage

At the point of capture, the primary defensive goal is data minimization: preventing raw sensitive signals from ever reaching an untrusted environment, rather than attempting to scrub identity from that signal after the fact. This section reviews four complementary families of collection-stage techniques.

3.1. On-Device and Edge Preprocessing

Performing inference directly on the capturing device (edge computing) removes the need to expose raw data to the cloud at all. Hybrid architectures split deep networks between device and cloud, transmitting only intermediate, task-specific representations rather than raw pixels or waveforms. Combined with noise injection at the split point, these architectures measurably reduce private-attribute leakage while preserving downstream task accuracy [15]. The design choice that matters most in practice is where the split occurs: splitting too early in the network leaves representations close to raw data (and therefore still recoverable via inversion, see Section 6); splitting too late reduces the communication savings that motivated edge computing in the first place. This trade-off is rarely made explicit in deployment guidance and represents a practical gap between the academic literature and production systems.

3.2. Obfuscation at Capture Time

Classical pixel-level operations—blurring, pixelization, and masking—remain widely deployed because they are cheap, well understood, and require no model training. Empirical studies show that their effectiveness is filter- and task-dependent: moderate blurring retains sufficient activity cues for human observers while degrading identity recognition, but the same blur radius that defeats a 2015-era face recognizer may be insuffi-

cient against a modern, more capable model trained specifically to recognize blurred faces [4,14]. This is the first concrete instance of the "adaptive adversary" problem introduced in Section 2 and revisited as an open challenge in Section 10: simple filtering offers only weak, non-quantifiable guarantees against a sufficiently motivated, well-resourced re-identification adversary.

3.3. Metadata Sanitization

EXIF, ID3, and similar embedded tag formats often contain GPS coordinates, device serial numbers, and precise timestamps. Failure to strip this metadata has been shown to leak precise location even when the visual or acoustic content itself has been carefully sanitized [3]. Because metadata sanitization is computationally negligible relative to visual/acoustic obfuscation or cryptographic techniques, it should be treated as a baseline, near-zero-cost complement applied in every pipeline rather than as an optional extra—yet in practice it is frequently overlooked precisely because it is unglamorous relative to, say, GAN-based de-identification.

3.4. Local Differential Privacy and Federated Collection

Local differential privacy (LDP) perturbs data on each device before transmission, using a randomized mechanism calibrated so that any single record's contribution to the aggregate is provably bounded, allowing useful population-level aggregation despite the added noise [16]. Federated collection extends this principle architecturally by avoiding centralization altogether: raw data remains on the device indefinitely, and only model updates (not data) ever leave it. This naturally motivates—and blurs the boundary with—the analysis-stage federated learning techniques discussed in Section 4; in practice, LDP and federated learning are frequently deployed together, with LDP bounding what any individual update can reveal and federated learning bounding what data ever needs to leave the device in the first place.

4. Privacy-Preserving Techniques at the Analysis Stage

The analysis stage—where data is mined, models are trained, and inferences are drawn—has attracted the most research attention of the three lifecycle stages, because it presents the sharpest tension between utility (the entire reason the data was collected) and exposure (the model or its outputs may now encode information about individual training records). This section reviews seven technique families operating at this stage.

4.1. Differential Privacy (DP) in Multimedia Analytics

Differential privacy provides a formal, composable guarantee that the presence or absence of any single record has a bounded effect on the distribution of a mechanism's output, typically enforced by injecting calibrated noise [17]. Formally, a randomized mechanism M satisfies (ε, δ)-differential privacy if for all datasets D and D′ differing in a single record and for all measurable output sets S, $\Pr[M(D) \in S] \leq e^{\varepsilon} \cdot \Pr[M(D') \in S] + \delta$, where ε bounds the worst-case privacy loss and δ allows a small probability of a stronger guarantee failing. In deep learning, this is operationalized via DP-SGD, which clips each per-example gradient to a bounded norm and adds calibrated Gaussian noise before the aggregated update is applied, together with a privacy accountant that tracks cumulative ε across training iterations [18]. Multimedia analytics face additional difficulties beyond those seen in tabular DP: images and audio are extremely high-dimensional, per-example gradients in convolutional and transformer architectures are noisy by nature even without added DP noise, and long training runs on large models interact with privacy budgets in ways that can force an uncomfortable choice between model quality and a defensible ε. Despite these difficulties, DP remains the principal mechanism for quantifiable, worst-case

privacy guarantees in model training and is often combined with federated or generative techniques rather than deployed alone [19].

4.2. Federated Learning (FL) for Distributed Analytics

Federated learning trains a shared model across distributed data holders (hospitals, personal devices) by exchanging model updates rather than raw data, following the original decentralized training formulation of McMahan et al. [20]. Medical imaging is the most mature application domain: frameworks have been demonstrated for histopathology whole-slide-image analysis [21], heterogeneous multi-institutional analysis across differing imaging protocols [22], and uncertainty-aware diagnostic imaging that additionally reports model confidence alongside predictions [23]. However, vanilla FL is not sufficient for strong privacy on its own: because model updates are computed directly from local data, they can leak substantial information about that data. Comprehensive analyses of deep learning privacy show that membership inference attacks can achieve up to 90% accuracy purely from shared updates, without any access to raw data [24], and dedicated gradient-inversion attacks can reconstruct near-pixel-perfect training images from a single shared gradient [13]. Consequently, FL is now understood as a necessary but insufficient building block: it must be combined with DP, secure aggregation, or both to provide meaningful protection [25], a conclusion that is now the consensus position in GDPR-oriented analyses of federated systems [26].

4.3. Homomorphic Encryption (HE) for Encrypted Computation

Homomorphic encryption allows arithmetic operations to be performed directly on ciphertexts, producing an encrypted result that, once decrypted, matches the result of performing the same operation on the plaintext. Early work demonstrated homomorphic JPEG encoding and image resizing [27], and subsequent schemes based on the CKKS approximate-arithmetic scheme extended this to brightness filtering and denoising operations while explicitly modeling spatial image structure [28,29]. Combining HE with data partitioning strategies reduces computational overhead relative to encrypting an entire image or video frame at once [30]. Persistent limitations, however, remain severe from a systems perspective: ciphertext expansion of one to three orders of magnitude relative to plaintext, the absence of efficient native support for non-linear operations (the activation functions ubiquitous in deep networks), and end-to-end latency that is currently unsuitable for anything beyond still-image or short-clip workloads [31]. This gap—the near-total absence of real-time, video-scale HE—is one of the most consequential open problems identified in Section 10.

4.4. Secure Multi-Party Computation (SMC)

Secure multi-party computation enables two or more parties to jointly compute a function over their private inputs (e.g., a face-similarity score) while revealing nothing beyond the function's output. SMC protocols based on techniques such as garbled circuits and secret sharing have been used for privacy-preserving face matching and image similarity search without either party disclosing its raw biometric template [32]. SMC shares with HE a fundamental tension between the strength of its cryptographic guarantee and its latency and scalability, which has motivated hybrid approaches that use SMC only for a small, security-critical sub-computation (such as a final similarity comparison) while performing feature extraction and indexing in the clear or under weaker protection.

4.5. Face and Visual De-Identification

Face de-identification is the most extensively studied multimedia-specific privacy technique, and its history illustrates the broader arms-race dynamic that motivates this review.

The foundational k-Same algorithm averages each face with its k−1 most visually similar faces in a reference set, guaranteeing a formal linkage probability of at most 1/k for a face-recognition adversary restricted to that reference set [33]. Deep-learning extensions such as k-Same-Net improve the visual realism of the averaged output while retaining the underlying k-anonymity guarantee [34]. Generative adversarial networks, following the original formulation of Goodfellow et al. [35], are now the dominant generative substrate for de-identification: attribute-transformation GANs replace real faces with entirely synthetic ones that preserve non-identifying attributes such as expression; disentangled-representation methods separate identity from other facial attributes, enabling personalized and even invertible de-identification for authorized parties [36]; two-step pipelines with explicit explainability components provide interpretable identity removal that can be audited [37]; and latent-code optimization methods anonymize entire datasets at once while preserving pose and expression distributions needed for downstream training [38]. Domain-specific variants include "digital masks" that preserve diagnostically relevant facial motion in telemedicine [39,40] and reversible ophthalmic anonymizers that retain clinically significant ocular signs while suppressing identity for general viewing [41]. Despite this rapid technical progress, broad surveys of the visual-privacy literature caution that filtering-based approaches alone—however visually convincing—cannot guarantee privacy against future, more capable recognition models, a concern that echoes the adaptive-adversary problem raised throughout this review [4,42].

4.6. Privacy-Preserving Action and Anomaly Recognition

A distinct line of work targets video analytics tasks (action recognition, anomaly detection) where the goal is to preserve behavioral signal while suppressing identity signal, rather than to remove faces outright. Approaches include extreme downsampling and motion-difference quantization that discard fine-grained identity cues while retaining coarse motion patterns [43]; adversarial training objectives that explicitly minimize mutual information between learned video features and private attributes such as identity or demographic group [44]; spatio-temporal architectures designed to resist identity leakage that accumulates across time, which is a distinct threat from single-frame leakage [45]; and lightweight "anonymizing adapters" that can be attached to large pretrained video models, which additionally have the side benefit of reducing demographic bias in the adapted model's predictions [46]. For city-scale surveillance specifically, the individually supported components discussed above—non-identifiable intermediate representations such as skeleton keypoints or optical flow fields rather than raw pixels, combined with HE and DP within an edge–cloud trust boundary—could in principle be layered into a single deployment pattern. However, our literature search did not identify a published system that implements and empirically evaluates this specific three-way combination; we therefore present it here as a composition suggested by the constituent techniques surveyed in this section, not as an established or independently validated architecture, and we flag scalable, empirically validated integration of these components as an open direction (Section 10).

4.7. Speech and Voice Privacy

Speaker anonymization alters an audio recording so that it cannot be linked back to the original speaker while preserving its linguistic content (i.e., what was said, not who said it). The VoicePrivacy Challenge established common datasets and evaluation protocols for this task, with baseline systems built on x-vector substitution—replacing the speaker-identity embedding while keeping the linguistic content embedding unchanged [47]. GAN-based approaches generate entirely artificial speaker embeddings rather than substituting

a real alternate speaker [48,49], while real-time anonymization systems are specifically optimized to preserve intelligibility and vocal timbre under low-latency constraints suitable for live calls [50]. Differentially private voice conversion has additionally been proposed for sharing sensitive health-related speech, such as recordings used to detect dementia biomarkers, while suppressing speaker identity under a formal privacy guarantee [51]. Robust evaluation frameworks such as VoicePM jointly assess both privacy (resistance to speaker re-identification) and utility (word-error rate, perceived naturalness), reflecting the same privacy–utility duality introduced in Section 2, and are increasingly viewed as a template that other modalities lack [52].

5. Privacy-Preserving Techniques at the Dissemination Stage

Once data, models, or derived statistics must be shared with external parties—researchers, regulators, the public, or commercial partners—the defensive challenge shifts from preventing exposure during processing to preventing re-identification or unauthorized secondary use after release, when the discloser typically has little or no further control over the data.

5.1. Anonymized Dataset Publishing

K-anonymity-based de-identification, discussed in Section 4, is often deployed as the final processing step immediately before public release. However, dedicated vulnerability studies show that face-obfuscated video can still leak identity through body shape, gait, clothing, or surrounding context that the obfuscation method did not target, underscoring that visual k-anonymity guarantees must be evaluated against a realistically capable, modern recognition adversary rather than the adversary the original technique was designed against [11].

5.2. Synthetic Data Generation

An increasingly popular alternative to publishing real (even if de-identified) data is releasing entirely synthetic data generated under a formal privacy constraint. DP-GAN [53], DP-CGAN [54], and PATE-GAN, which uses a teacher–student knowledge-distillation architecture to bound privacy loss [55], all provide formal (ε, δ)-DP guarantees on the generative model rather than on individual output samples directly. Domain-specific extensions for medical imaging adapt these mechanisms to clinical data requirements, such as preserving diagnostically relevant texture statistics [56]. A consistent finding across this literature is that tightening the privacy budget improves resistance to membership-inference attacks (Section 6) but reduces sample fidelity—a fundamental trade-off that becomes markedly more severe in high-dimensional settings like images than in the low-dimensional tabular settings where DP-synthetic-data methods were first developed [55].

5.3. Blockchain-Enabled Controlled Dissemination

Blockchain provides cryptographic, decentralized governance over data access without relying on a single trusted intermediary. In vehicular networks, pseudonym-based schemes allow dashcam imagery to be shared under rotating pseudonyms, enabling after-the-fact accountability (e.g., for insurance or law enforcement purposes) while preventing routine identity disclosure during normal operation [57]. Decentralized data-transaction frameworks address the single-point-of-failure and unilateral-control weaknesses inherent to centralized data-sharing platforms [58]. Blockchain-based intellectual-property authentication supports traceable provenance for shared media assets [59], and hybrid AI-blockchain architectures have been proposed specifically for critical infrastructure, where both data integrity and confidentiality must be jointly assured [60]. Systematic reviews of this space consistently emphasize, however, that blockchain alone provides access governance and auditability but not confidentiality; it must be combined with encryption and off-chain

storage of the sensitive payload itself to deliver a complete confidentiality guarantee, since data written directly to a public or even permissioned ledger is not inherently private [61].

5.4. Access Control, Watermarking, and Fair Exchange

Dissemination pipelines increasingly combine the mechanisms above with digital watermarking for provenance tracing (identifying the source of a leak after the fact), attribute-based access control for fine-grained selective disclosure (e.g., releasing de-identified imagery to researchers but full-resolution imagery only to law enforcement under warrant), and fair-exchange protocols that ensure neither party can unilaterally defect during a data transaction. These mechanisms provide auditability and accountability that purely technical anonymization cannot offer once data has left the originator's direct operational control—a distinction that is often underemphasized in purely cryptographic treatments of dissemination privacy.

6. Privacy Attacks That Motivate These Defenses

This section reviews, in more technical depth than the motivating remarks in Sections 3–5, the four attack families that most directly shape the design of the defenses surveyed in this review.

6.1. Re-Identification and De-Anonymization

Re-identification attacks recover the identity of an individual from ostensibly anonymized content by exploiting auxiliary information that the anonymization procedure did not account for—for instance, correlating de-identified gait or clothing with a separate, identified dataset, or exploiting recognition capability from a model released after the anonymization was designed [11]. The defining feature of this attack family, distinguishing it from model-centric attacks, is that it operates purely on released outputs (images, video, and aggregate statistics) and requires no access whatsoever to the model or system that produced the anonymization.

6.2. Membership Inference Attacks (MIAs)

Membership inference attacks determine whether a specific record was present in a model's training set, typically by exploiting the fact that models tend to behave more confidently or produce systematically different outputs on data they were trained on versus data they were not. Text-to-image generative models are especially vulnerable to this class of attack, with demonstrated attacks achieving near-certainty in distinguishing training members from non-members [62]. Diffusion models are similarly vulnerable via attacks that exploit the statistical structure of the initial noise sample used during generation [63] or that exploit perceptual-loss signals computed during the generation process [64]. These attacks directly motivate the DP-constrained training and generation techniques surveyed in Section 5, since a formal DP guarantee on the training procedure provably bounds the maximum success rate any membership-inference adversary can achieve, regardless of the specific attack strategy used.

6.3. Gradient and Model Inversion

In federated settings, a malicious participant or a compromised central aggregator can attempt to reconstruct raw training images directly from the model updates shared during training. The Deep Leakage from Gradients attack demonstrated that this reconstruction can be near-pixel-accurate for images and token-accurate for text, using only a single shared gradient and no auxiliary information about the underlying data [13]. This risk is compounded by model-inversion attacks in the deployed (post-training) setting, where an adversary with only query access to a trained model—not access to gradients

during training—can exploit confidence scores returned alongside predictions to reconstruct recognizable training-set face images [12]. Comprehensive privacy analyses of deep learning models confirm that inversion-style attacks combined with membership inference can achieve accuracy as high as 90% from model updates alone [24]. Together, these results justify combining federated learning with secure aggregation, homomorphic encryption, or differential privacy, as recommended throughout Section 4, rather than treating federation itself as a sufficient privacy mechanism.

6.4. Metadata and Location Inference

Embedded metadata can directly expose precise GPS coordinates and timestamps even when the associated visual or audio content is otherwise well protected [3,64]. At the aggregate level, membership-inference attacks against released, ostensibly noise-protected location statistics demonstrate that even carefully aggregated and perturbed data can leak individual-level contributions under a sufficiently informed adversary, echoing the same fundamental tension between aggregate utility and individual privacy seen throughout this review [65].

7. Comparative Taxonomy

Table 2 provides a multi-stage overview of privacy-preserving paradigms mapping onto lifecycle dimensions, modality constraints, and analytical utility penalties, synthesizing the technique families discussed in detail in Sections 3–5.

Table 2. Comparative taxonomy of privacy-preserving techniques across the multimedia data lifecycle, indicating modality applicability, core mechanism, and typical utility impact.

Technique Family	Lifecycle Stage	Modality	Core Mechanism	Utility Impact	Computational Cost/Latency
On-device/edge preprocessing	Collection	Image, Video	Local extraction; minimal data transfer	Low–moderate	Not separately quantified in the reviewed literature
Visual/acoustic obfuscation	Collection, Analysis	Image, Video, Audio	Blurring, pixelization, masking	Moderate–high	Not separately quantified in the reviewed literature
Metadata sanitization	Collection	All	Strip/generalize EXIF, geotags	Negligible	Negligible (Section 3)
Local differential privacy	Collection, Analysis	Sensor streams, Features	Calibrated local noise injection	Moderate	Not separately quantified in the reviewed literature
Federated learning	Analysis	Image, Video (medical)	Decentralized training; share updates	Low–moderate	Not separately quantified in the reviewed literature
Homomorphic encryption	Analysis	Image (video)	Computation on ciphertext (CKKS)	Near-lossless	High (Section 4)
Secure multi-party computation	Analysis	Image (biometric)	Joint computation without raw inputs	High per Figure 3; not independently quantified in Section 4 text	High latency (Section 4)
k-anonymity/GAN de-identification	Analysis, Dissemination	Face image, Video	Averaging or generative substitution	Moderate	Not separately quantified in the reviewed literature
Privacy-preserving action/anomaly	Analysis	Video	Non-identifiable intermediate targets	Moderate	Not separately quantified in the reviewed literature
Voice/speaker anonymization	Analysis, Dissemination	Audio	Speaker-embedding substitution	Moderate	Not separately quantified in the reviewed literature
DP synthetic data generation	Dissemination	Image, text-to-image	DP-constrained generative models	Moderate–high	Not separately quantified in the reviewed literature
Blockchain-based sharing	Dissemination	Image, Video, Mixed	Pseudonymity, access control, provenance	Not separately quantified in the reviewed literature	Adds latency (Section 5)

DP = Differential Privacy; GAN = Generative Adversarial Network; CKKS = Homomorphic encryption scheme.

Figure 3 visualizes the privacy–utility trade-off summarized qualitatively in Table 2 as a two-dimensional landscape, with bubble size representing typical computational or latency cost. The coordinates and bubble sizes in this figure are our qualitative synthesis of the relative positioning of each technique family as characterized in the literature reviewed in Sections 3–5; they are not derived from a numerical scoring rubric or a reproducible measurement protocol and should be read as an approximate, expert-judgment visualization of relative standing rather than as precise measurements. Two patterns are apparent. First, cryptographic techniques (homomorphic encryption, secure multi-party computation) occupy the high-privacy, high-cost region of the landscape, consistent with the latency limitations discussed in Section 4. Second, the lowest-cost collection-stage techniques (metadata sanitization, on-device preprocessing) cluster toward high utility retention but comparatively modest privacy strength on their own, consistent with their role as necessary but individually insufficient first-line defenses rather than complete solutions. This pattern does not extend to every collection-stage family: visual/acoustic obfuscation, also assigned to the collection stage, sits toward the lower end of the utility-retention axis, consistent with Table 2's classification of its utility impact as moderate-to-high rather than negligible. Similarly, Figure 3 places secure multi-party computation toward the higher end of utility retention even though Table 2 separately characterizes it as high-latency; utility retention (fidelity of the computed output) and computational or latency cost are related but distinct axes in our taxonomy, and a technique such as SMC can preserve output fidelity closely while remaining expensive to run, which is why cost is encoded separately via bubble size rather than folded into the utility-retention axis.

Figure 4 further breaks down technique coverage by modality. A clear gap emerges: audio-specific techniques (voice/speaker anonymization) do not extend to image or video, and conversely, most video-oriented techniques (action/anomaly recognition, k-anonymity/GAN de-identification) have only partial or no established audio counterpart. Very few technique families—chiefly visual/acoustic obfuscation and blockchain-based sharing—have been demonstrated across all three core modalities (image, video, and audio), which suggests that genuinely modality-general privacy-preserving multimedia frameworks remain an open, underexplored direction rather than a solved problem, a point we develop further in Sections 8 and 10.

"Established" indicates multiple independent studies demonstrating the technique on that modality; "partial" indicates emerging or single-study evidence; a blank cell indicates no identified application in the reviewed literature. This classification reflects our qualitative assessment of the literature surveyed in Sections 3–6 rather than a numerical threshold or a formally coded rubric; independence of evidence within a cell was judged narratively (e.g., studies from the same research group or reusing the same dataset were weighted as a single source of evidence rather than as independent confirmations). Supplementary Table S1 traces each cell to the specific reference numbers cited in the corresponding section and flags the small number of cells for which our audit did not find a specific supporting citation in the current text.

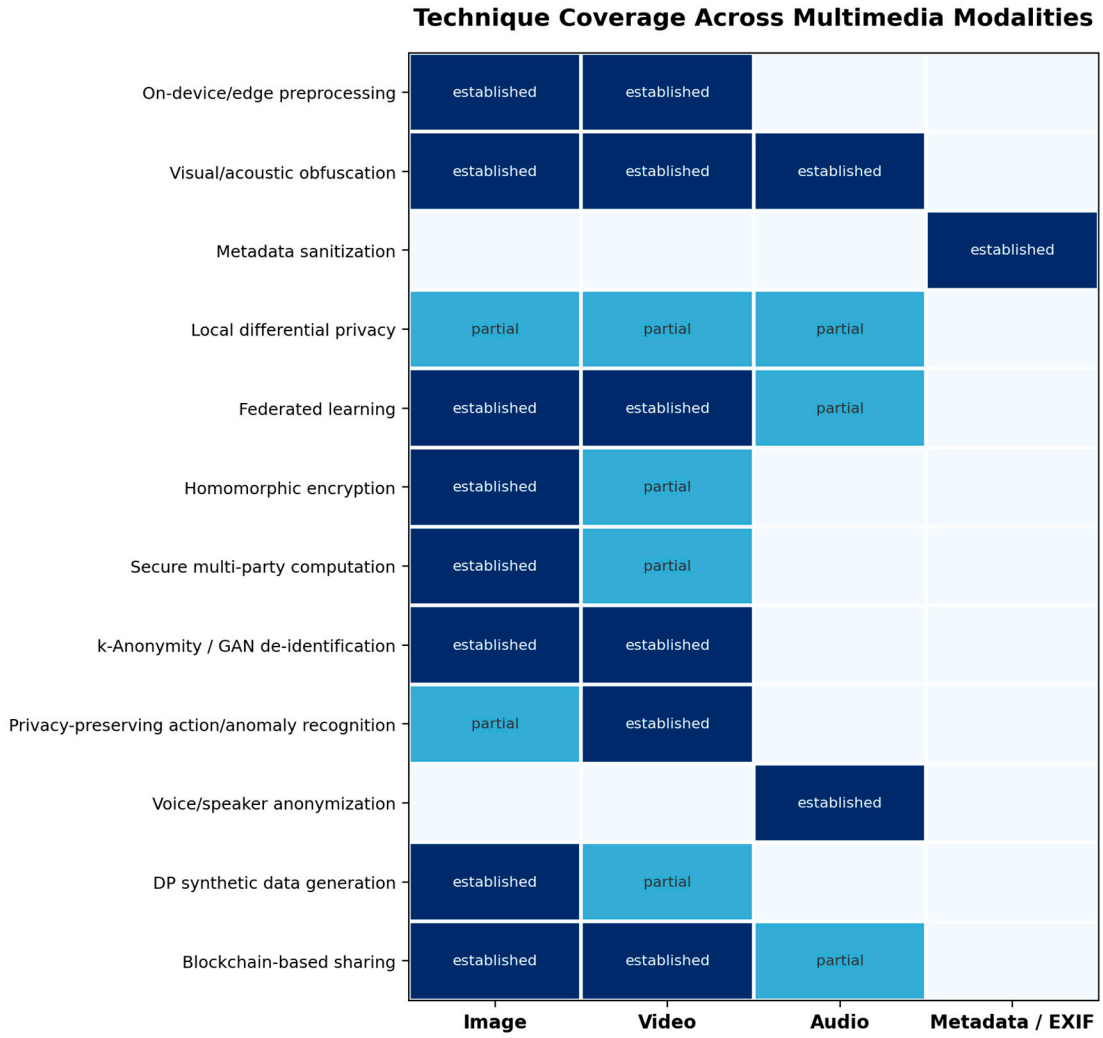


Figure 4. Coverage of each technique family across image, video, audio, and metadata modalities (rating definitions and evidence basis in the text and Supplementary Table S1).

8. Discussion: Cross-Cutting Themes and Tensions

Having reviewed each lifecycle stage and its motivating attacks in turn, we now step back and synthesize four themes that recur across, rather than within, Sections 3–7.

A pattern repeated across nearly every technique family surveyed is that the mechanism, deployed alone, is demonstrably insufficient against a realistic adversary, and the literature's own trajectory has been to layer mechanisms rather than to seek a single dominant solution. Federated learning requires DP or secure aggregation (Section 4); blockchain requires encryption and off-chain storage (Section 5); and visual obfuscation requires evaluation against and often augmentation by learned de-identification (Sections 4 and 5). This suggests that the appropriate unit of comparison for future work may increasingly be validated composite pipelines rather than individual point techniques, and that reviews and benchmarks (including this one) should be read with this composability requirement in mind.

Several results surveyed in this review (Sections 3–5) show a defense that was adequate against the recognition capability available at publication time being defeated by a subsequently released, more capable model. Yet the great majority of evaluation protocols in the surveyed literature benchmark against a fixed, contemporaneous adversary rather than attempting to model or bound performance against a future, more capable one. We view this as a methodological gap rather than a purely technical one: it may be addressed

by adopting evaluation practices from adjacent fields (e.g., adversarial robustness research), which routinely evaluate against adaptive, worst-case, and future-anticipated threat models rather than fixed benchmarks.

Differential privacy's ε is a rigorous, composable, and widely accepted guarantee, but it is not directly interpretable in terms that a data subject, clinician, or downstream task designer would find meaningful ("how much can this ε leak about me, in terms I understand?"). Frameworks such as VoicePM (Section 4) that jointly report a privacy metric and a task-relevant utility metric (word-error rate, naturalness) are, in our reading of this literature, the exception rather than the norm, and no comparably mature joint metric yet exists for image or video modalities. We consider this the single most actionable open problem identified in this review and expand on it in Section 10.

As Figure 4 makes visually explicit, audio-privacy and video-privacy research communities have developed largely in parallel, with limited cross-pollination of technique or evaluation methodology. Voice anonymization has a mature, standardized, community-wide benchmark (the VoicePrivacy Challenge); no equivalently mature, standardized, community-wide benchmark yet exists for video de-identification or action-recognition privacy, despite video representing a substantially larger share of sensitive multimedia traffic in practice.

A fifth, more self-referential theme emerged only once we audited our own citations while assembling Supplementary Table S1 (Section 7): the strength of evidence behind a given technique–modality pairing varies considerably even within a single technique family, and this variation is not visible from Figure 4 or Table 2 alone. Some pairings we and the figure both describe qualitatively as at least partially supported rest on a single, non-modality-specific study (on-device/edge preprocessing, cited only by [15] for both image and video); others rest on no dedicated citation in the literature we reviewed at all (audio coverage for visual/acoustic obfuscation and for blockchain-based sharing; video coverage for federated learning, secure multi-party computation, and DP synthetic data generation; image coverage for privacy-preserving action/anomaly recognition). We view this less as a flaw specific to those technique families and more as a general property of a fast-moving field, where extending a mechanism to a new modality is often assumed straightforward by practitioners and therefore left unpublished as a dedicated, citable study. We flag it explicitly here, rather than only in Supplementary Table S1, because it bears directly on how the comparative taxonomy in Section 7 should be read: modality-coverage claims are best treated as directional evidence of where research attention has concentrated, not as an exhaustive or uniformly reliable map of technical feasibility. Closing these specific, identified evidence gaps is a concrete and comparatively low-cost contribution the community could make in the near term, distinct from the more open-ended research challenges we turn to in Section 10.

9. Application Domains

This section discusses five domains in which the lifecycle-stage techniques reviewed in Sections 3–5 are deployed jointly, illustrating how the abstractions of the preceding sections combine in practice.

9.1. Smart Video Surveillance and Smart Cities

City-scale camera networks require reconciling public-safety objectives (crowd counting, incident and anomaly detection) with bystander privacy for individuals who did not consent to being recorded and often are not even aware of the camera's presence. Layered architectures combining edge-based de-identification (Sections 3 and 4), encrypted or aggregated analytics, and access-controlled retention policies are the most common design

pattern in the literature we reviewed, reflecting the composability theme of Section 8 [24]. In practice, the retention-policy dimension is at least as consequential for realized privacy risk as the choice of technical de-identification mechanism, since footage retained indefinitely under weak governance remains a latent risk regardless of how it was initially processed.

9.2. Telemedicine and Medical Imaging

Remote consultation video, dermatological and ophthalmic imaging, and digital histopathology all require sharing visual data across institutional boundaries while complying with health-data regulations. Federated learning combined with DP and secure aggregation, together with de-identification methods that explicitly preserve clinically significant features such as facial motion or ocular signs (Section 4), exemplify the domain-specific adaptations this field has produced [21–23,40,41]. Medical imaging is, in our assessment, the single most mature application domain surveyed in this review, likely because the regulatory stakes (HIPAA, GDPR Article 9) and the clinical value of the data have jointly justified sustained, well-funded research investment relative to other domains.

9.3. Social Media and Content Platforms

Large-scale privacy risk on social media and content-sharing platforms arises from embedded metadata, automatic face recognition applied to uploaded content, and downstream model training—including the training of generative models—on that user-uploaded content, often under terms of service that provide only nominal user awareness of this secondary use. Solutions in the surveyed literature include metadata sanitization on upload, opt-in de-identification for facial content, and DP- or MIA-resistant training pipelines for any models trained on the resulting corpus [3,62]. This domain is distinguished from the others reviewed by the sheer scale and heterogeneity of the data involved, which limits the applicability of computationally expensive techniques such as HE or SMC (Section 4) and instead favors lighter-weight collection-stage and DP-based analysis-stage mechanisms.

9.4. Connected and Autonomous Vehicles

Dashcam and roadside imagery shared for cooperative perception (e.g., vehicles pooling sensor data to improve collective situational awareness) can inadvertently capture identifiable bystanders and license plates well beyond the vehicle's own occupants. Pseudonym-based blockchain frameworks enable accountable, privacy-preserving exchange of this imagery, allowing after-the-fact accountability without routine identity disclosure during normal cooperative-perception operation [57]. Incentive-based strategies have also been proposed to preserve privacy in vehicular cloud settings [66]. This domain is notable for combining, within a single application, collection-stage obfuscation of bystanders, analysis-stage federated perception models shared across vehicles, and dissemination-stage blockchain governance—arguably the most complete real-world instance of the full lifecycle-composability argument made in Section 8.

9.5. Voice Assistants and IoT Audio

Always-listening consumer devices raise continuous speaker-privacy concerns distinct from the episodic capture typical of cameras, since the device is, by design, always positioned to capture audio. Speaker anonymization and differentially private voice conversion (Section 4) allow the resulting audio to be used for service improvement and even health monitoring (e.g., detecting changes in speech patterns relevant to cognitive health) without persistent identity exposure to the service provider [48–51]. The always-on nature of this domain makes on-device (collection-stage) processing particularly attractive relative to cloud-based analysis, since it avoids continuous transmission of raw, unprotected audio entirely.

10. Open Challenges and Future Research Directions

Synthesizing the technical review (Sections 3–6) and cross-cutting discussion (Section 8), we identify seven concrete open challenges that we believe should guide the next phase of research in privacy-preserving multimedia systems.

10.1. Scalable Encrypted and Federated Processing for Video

Most HE and SMC demonstrations reviewed in Section 4 remain limited to still images or short clips; achieving real-time, video-scale encrypted analytics is a priority that will likely require hardware-accelerated cryptographic primitives (e.g., GPU- or FPGA-accelerated CKKS) and approximate cryptographic schemes that deliberately trade a small amount of numerical precision for substantially improved throughput [31].

10.2. Standardized, Modality-Aware Privacy–Utility Metrics

As argued in Section 8, DP's ε is not directly interpretable in terms of perceptual or re-identification risk for a non-specialist stakeholder. The field needs metrics that jointly capture statistical privacy guarantees and perceptual or task-recognition-based utility loss in a single, standardized reporting format, building on the emerging joint privacy/utility evaluation template established by VoicePM for speech [52] and extending an analogous standard to image and video modalities, where no comparably mature equivalent currently exists.

10.3. Robustness Against Adaptive, Foundation-Model-Era Adversaries

De-identification methods evaluated only against contemporaneously available recognition systems can be, and repeatedly have been, defeated by subsequently released, more capable models—large-scale foundation-model embeddings or diffusion-based reconstruction methods not available when the original defense was published [4,11]. Evaluation protocols must anticipate this drift explicitly, for instance by reporting a defense's expected "shelf life" against a projected future adversary rather than only its performance against present-day baselines, an evaluation norm largely absent from the literature we reviewed.

10.4. Principled Composition of Hybrid Defenses

Combining federated learning, DP, HE, and secure aggregation is increasingly common in deployed systems (Section 8), but the compounded utility loss and the interaction between multiple simultaneously tracked privacy budgets are not well characterized for multimedia specifically, as opposed to the tabular-data setting where composition theorems are comparatively mature [25,26]. A principled accounting framework for multimedia-specific composed privacy budgets—analogous to the moments accountant developed for DP-SGD—remains an open and, in our view, tractable near-term research target.

10.5. Privacy–Fairness Interactions

Anonymization techniques can affect protected demographic attributes unevenly, for instance, if a de-identification model performs less faithfully on underrepresented skin tones or accents. Recent work, however, shows that carefully designed anonymization can simultaneously reduce both privacy leakage and demographic bias in the resulting model, suggesting that co-optimizing these two objectives—rather than treating fairness as an afterthought to be checked once a privacy technique is finalized—is achievable and should become standard practice [46].

10.6. Regulatory Harmonization and Auditability

GDPR, HIPAA, CCPA/CPRA, and PIPL impose related but non-identical obligations on biometric and health-related multimedia data, creating compliance complexity for

any organization operating across jurisdictions [6]. Technical systems would benefit substantially from auditable, certifiable privacy-by-design tooling capable of demonstrating compliance against multiple regulatory frameworks simultaneously, rather than requiring bespoke compliance engineering for each jurisdiction a system is deployed in.

10.7. Usable Privacy for Data Controllers and Subjects

Most techniques surveyed in this review are evaluated exclusively on technical metrics—privacy budgets, re-identification rates, and downstream task accuracy. Relatively little of the literature we reviewed assesses whether data subjects and the operational staff deploying these systems actually understand the protection being applied, or can meaningfully exercise informed consent regarding it [19]. We view human–computer interaction research into usable privacy explanation and consent interfaces for multimedia-specific mechanisms as a substantially underexplored complement to the primarily cryptographic and statistical contributions reviewed in Sections 3–6.

11. Conclusions

Multimedia data poses privacy challenges qualitatively different from those of structured data: identity is encoded redundantly and implicitly across pixels, waveforms, and metadata, and continuing deep learning progress steadily erodes earlier anonymization techniques that were adequate when first proposed. This review organized the body of privacy-preserving techniques along the multimedia data lifecycle—collection, analysis, and dissemination—covering on-device minimization and metadata sanitization at capture; differential privacy, federated learning, homomorphic encryption, secure multi-party computation, and learned de-identification during analysis; and anonymized publishing, DP-constrained synthetic data generation, and blockchain-based controlled sharing at dissemination. We further related these defenses directly to the attacks that motivate them—re-identification, membership inference, gradient and model inversion, and metadata/location inference—and synthesized both a comparative taxonomy and a set of cross-cutting themes (Section 8) that, in our assessment, are currently more consequential open problems than the invention of further individual point techniques.

No single technique offers complete protection; each carries a distinct privacy–utility trade-off and computational cost, and robust deployments increasingly combine several mechanisms across, and even within, individual lifecycle stages. As multimedia datasets continue to grow and recognition models continue to improve, the central open problem is progressively less about inventing new individual mechanisms and increasingly about composing, scaling, standardizing the evaluation of, and auditing these mechanisms coherently—an agenda spanning cryptography, statistics, machine learning, human–computer interaction, and law.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/multimedia2040018/s1, Table S1: Evidence Basis for Figure 4.

Author Contributions: Conceptualization, H.M. and M.S.A.N.; methodology, H.M. and M.S.A.N.; formal analysis H.M. and Z.Z.; investigation, H.M., Z.Z. and M.S.A.N.; writing—original draft preparation, H.M., Z.Z. and M.S.A.N.; visualization, H.M., Z.Z. and M.S.A.N.; supervision, H.M.; project administration H.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: During preparation of this work, the authors used a generative AI tool (Claude Sonnet 5.5) to assist with drafting narrative text and to generate the visual summaries in Figures 1–4. AI-assisted content, including figures, technical claims, and the reference list, has subsequently been reviewed and edited by the authors.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Cisco. *Cisco Annual Internet Report (2018–2023)*; White Paper; Cisco: San Jose, CA, USA, 2020.
2. Cao, Q.; Shen, L.; Xie, W.; Parkhi, O.M.; Zisserman, A. VGGFace2: A Dataset for Recognising Faces across Pose and Age. In Proceedings of the 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), Xi'an, China, 15–19 May 2018; pp. 67–74.
3. Donath, J. *The Social Machine: Designs for Living Online*; MIT Press: Cambridge, MA, USA, 2014.
4. Padilla-López, J.R.; Chaaraoui, A.A.; Flórez-Revuelta, F. Visual Privacy Protection Methods: A Survey. *Expert Syst. Appl.* **2015**, *42*, 4177–4195. [CrossRef]
5. Tolosana, R.; Vera-Rodriguez, R.; Fierrez, J.; Morales, A.; Ortega-Garcia, J. DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection. *Inf. Fusion* **2020**, *64*, 131–148. [CrossRef]
6. European Commission. Regulation (EU) 2016/679 (GDPR). *Off. J. Eur. Union* **2016**, *L119*, 1–88.
7. Zhao, R.; Zhang, Y.; Wang, T.; Wen, W.; Xiang, Y.; Cao, X. Visual Content Privacy Protection: A Survey. *ACM Comput. Surv.* **2025**, *57*, 1–36. [CrossRef]
8. Wen, W.; Yuan, Z.; Zhang, Y.; Wang, T.; Xiao, X.; Zhao, R.; Fang, Y. Image Privacy Protection: A Survey. *ACM Comput. Surv.* **2026**, *58*, 1–36. [CrossRef]
9. Madhusudhanan, S.; Jose, A.C. Privacy Preservation Techniques through Data Lifecycle: A Comprehensive Literature Survey. *Comput. Secur.* **2025**, *155*, 104473. [CrossRef]
10. Jain, A.K.; Ross, A.; Nandakumar, K. *Introduction to Biometrics*; Springer: Berlin/Heidelberg, Germany, 2011.
11. Kikuchi, H.; Miyoshi, S.; Hernandez-Matamoros, A. A Vulnerability in Video Anonymization—Privacy Disclosure from Face-Obfuscated Video. In Proceedings of the 19th Annual International Conference on Privacy, Security and Trust (PST), Fredericton, NB, Canada, 22–25 August 2022; pp. 1–10.
12. Fredrikson, M.; Jha, S.; Ristenpart, T. Model Inversion Attacks that Exploit Confidence Information and Basic Countermeasures. In Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security (CCS), Denver, CO, USA, 12–16 October 2015; pp. 1322–1333.
13. Zhu, L.; Liu, Z.; Han, S. Deep Leakage from Gradients. In *Advances in Neural Information Processing Systems (NeurIPS 32)*; Neural Information Processing Systems Foundation, Inc.: San Diego, CA, USA, 2019; pp. 14747–14756.
14. Senior, A.; Pankanti, S.; Hampapur, A.; Brown, L.; Tian, Y.L.; Connell, J.; Bolle, R.M. Privacy protection in a smart surveillance system. In *Protecting Privacy in Video Surveillance*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 35–51.
15. Osia, S.A.; Shamsabadi, A.S.; Sajadmanesh, S.; Taheri, A.; Katevas, K.; Rabiee, H.R.; Lane, N.D.; Haddadi, H. A Hybrid Deep Learning Architecture for Privacy-Preserving Mobile Analytics. *IEEE Internet Things J.* **2020**, *7*, 5949–5960. [CrossRef]
16. Wang, T.; Hu, Z. Real-Time Stream Statistics via Local Differential Privacy in Mobile Crowdsensing. In *Proceedings of the International Conference on Mobile Multimedia Communications*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 432–445.
17. Dwork, C.; Roth, A. The Algorithmic Foundations of Differential Privacy. *Found. Trends Theor. Comput. Sci.* **2014**, *9*, 211–407. [CrossRef]
18. Abadi, M.; Chu, A.; Goodfellow, I.; McMahan, H.B.; Mironov, I.; Talwar, K.; Zhang, L. Deep Learning with Differential Privacy. In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS), Vienna, Austria, 24–28 October 2016; pp. 308–318.
19. Drechsler, J.; Bailie, J. *The Complexities of Differential Privacy for Survey Data*; NBER Working Paper No. 32905; National Bureau of Economic Research: Cambridge, MA, USA, 2024.
20. McMahan, H.B.; Moore, E.; Ramage, D.; Hampson, S.; Agüera y Arcas, B. Communication-Efficient Learning of Deep Networks from Decentralized Data. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), Ft. Lauderdale, FL, USA, 20–22 April 2017; pp. 1273–1282.
21. Lin, Y.; Wang, H.; Li, W.; Shen, J. Federated Learning with Hyper-Network—A Case Study on Whole Slide Image Analysis. *Sci. Rep.* **2023**, *13*, 1693. [CrossRef] [PubMed]

22. Zhou, J.; Zhou, L.; Wang, D.; Xu, X.; Li, H.; Chu, Y.; Han, W.; Gao, X. Personalized and Privacy-Preserving Federated Heterogeneous Medical Image Analysis with PPPML-HMI. *arXiv* **2023**, arXiv:2302.11571.
23. Koutsoubis, N.; Yilmaz, Y.; Ramachandran, R.P.; Schabath, M.; Rasool, G. Privacy-Preserving Federated Learning in Medical Imaging with Uncertainty Estimation. *arXiv* **2024**, arXiv:2406.12815.
24. Nasr, M.; Shokri, R.; Houmansadr, A. Comprehensive Privacy Analysis of Deep Learning. In Proceedings of the IEEE Symposium on Security and Privacy (SP), San Francisco, CA, USA, 21–23 May 2018; pp. 1–15.
25. Riedel, P.; von Schwerin, R.; Schaudt, D.; Hafner, A.; Späte, C. ResNetFed: Federated Deep Learning Architecture for Privacy-Preserving Pneumonia Detection from COVID-19 Chest Radiographs. *J. Healthc. Inform. Res.* **2023**, *7*, 203–224. [CrossRef] [PubMed]
26. Truong, N.; Sun, K.; Wang, S.; Guitton, F.; Guo, Y. Privacy Preservation in Federated Learning: An Insightful Survey from the GDPR Perspective. *arXiv* **2020**, arXiv:2011.05411.
27. Fu, W.; Lin, R.; Inge, D. Fully Homomorphic Image Processing. *arXiv* **2018**, arXiv:1810.03249.
28. Ziad, M.T.I.; Alanwar, A.; Alzantot, M.; Srivastava, M. CryptoImg: Privacy-Preserving Processing over Encrypted Images. In Proceedings of the IEEE Conference on Communications and Network Security (CNS), Philadelphia, PA, USA, 17–19 October 2016; pp. 570–575.
29. Cheon, J.H.; Kim, A.; Kim, M.; Song, Y.S. Homomorphic Encryption for Arithmetic of Approximate Numbers. In *Advances in Cryptology—ASIACRYPT 2017*; Springer: Cham, Switzerland, 2017; pp. 409–437.
30. Li, P.; Cai, D.; Yang, J.; Qian, W.; Wang, C.; Li, X.; Yang, S. Efficient and Privacy-Preserving Image Classification Using Homomorphic Encryption and Chunk-Based Convolutional Neural Network. *J. Cloud Comput.* **2023**, *12*, 78. [CrossRef]
31. Gentry, C. A Fully Homomorphic Encryption Scheme. Ph.D. Thesis, Stanford University, Stanford, CA, USA, 2009.
32. Kaaniche, N.; Laurent, M. Cloud-Based Privacy Preserving Image Storage, Sharing and Search. *arXiv* **2014**, arXiv:1410.6593.
33. Newton, E.M.; Sweeney, L.; Malin, B. Preserving Privacy by De-Identifying Face Images. *IEEE Trans. Knowl. Data Eng.* **2005**, *17*, 232–243. [CrossRef]
34. Meden, B.; Emeršič, Ž.; Štruc, V.; Peer, P. k-Same-Net: K-Anonymity with Generative Deep Neural Networks for Face Deidentification. *Entropy* **2018**, *20*, 60. [CrossRef] [PubMed]
35. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. In *Advances in Neural Information Processing Systems (NeurIPS 27)*; Neural Information Processing Systems Foundation, Inc.: San Diego, CA, USA, 2014; pp. 2672–2680.
36. Cao, J.; Liu, B.; Wen, Y.; Xie, R.; Song, L. Personalized and Invertible Face De-Identification by Disentangled Identity Information Manipulation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 11–17 October 2021; pp. 3314–3342.
37. Wen, Y.; Liu, B.; Cao, J.; Xie, R.; Song, L. Divide and Conquer: A Two-Step Method for High Quality Face De-Identification with Model Explainability. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 1–6 October 2023; pp. 5148–5157.
38. Barattin, S.; Tzelepis, C.; Patras, I.; Sebe, N. Attribute-Preserving Face Dataset Anonymization via Latent Code Optimization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 8001–8010.
39. Yang, Y.; Lyu, J.; Wang, R.; Wen, Q.; Zhao, L.; Chen, W.; Bi, S.; Meng, J.; Mao, K.; Xiao, Y.; et al. A Digital Mask to Safeguard Patient Privacy. *Nat. Med.* **2022**, *28*, 1883–1892. [CrossRef] [PubMed]
40. Zhu, B.; Zhang, C.; Sui, Y.; Li, L. FaceMotionPreserve: A Generative Approach for Facial De-Identification and Medical Information Preservation. *Sci. Rep.* **2024**, *14*, 17275. [CrossRef] [PubMed]
41. Tian, Y.; Zhou, M.; Chen, Y.; Li, F.; Qi, L.; Wang, S.; Xu, X.; Yu, Y.; Xu, S.; Lei, C.; et al. ROFI: A Deep Learning-Based Ophthalmic Sign-Preserving and Reversible Patient Face Anonymizer. *npj Digit. Med.* **2025**, *8*, 206. [CrossRef] [PubMed]
42. Ribaric, S.; Ariyaeeinia, A.; Pavesic, N. De-Identification for Privacy Protection in Multimedia Content: A Survey. *Signal Process. Image Commun.* **2016**, *47*, 131–151. [CrossRef]
43. Chen, X.; Wang, L. Privacy-Preserving Action Recognition: A Survey. In *Pattern Recognition and Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2024; pp. 201–215.
44. Fioresi, J.; Dave, I.R.; Shah, M. Privacy Beyond Pixels: Latent Anonymization for Privacy-Preserving Video Understanding. *arXiv* **2025**, arXiv:2511.08666.
45. Liu, Y.; Liu, S.; Zhu, X.; Li, J.; Yang, H.; Teng, L.; Guo, J.; Wang, Y.; Yang, D.; Liu, J. Privacy-Preserving Video Anomaly Detection: A Survey. *arXiv* **2024**, arXiv:2411.14565.
46. Cai, Z.; Gao, Z.; Planche, B.; Zheng, M.; Asif, M.S.; Wu, Z. Disguise without disruption: Utility-preserving face de-identification. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 20–27 February 2024; Volume 38, pp. 918–926.

47. Turner, H.; Lovisotto, G.; Martinovic, I. Speaker Anonymization with Distribution-Preserving X-Vector Generation for the VoicePrivacy Challenge 2020. *arXiv* **2020**, arXiv:2010.13457.
48. Meyer, S.; Tilli, P.; Denisov, P.; Lux, F.; Koch, J.; Vu, N.T. Anonymizing Speech with Generative Adversarial Networks to Preserve Speaker Privacy. In Proceedings of the IEEE Spoken Language Technology Workshop (SLT), Doha, Qatar, 9–12 January 2023; pp. 912–919.
49. Jafari, A.; Al-Mousa, A.; Jafar, I. Speaker Anonymization Using Generative Adversarial Networks. *J. Intell. Fuzzy Syst.* **2023**, *45*, 3345–3359. [CrossRef]
50. Deng, J.; Teng, F.; Chen, Y.; Chen, X.; Wang, Z.; Xu, W. V-Cloak: Intelligibility-, Naturalness- & Timbre-Preserving Real-Time Voice Anonymization. *arXiv* **2022**, arXiv:2210.15140.
51. Dawalatabad, N.; Ahangaran, M.; Karjadi, C.; Kourtis, L.; Joung, P.; Kolachalama, V.B.; Au, R.; Glass, J. Differential Privacy Preserving Voice Conversion for Audio Health Data. *Alzheimer's Dement.* **2025**, *21*, e091280.
52. Zhang, S.; Li, Z.; Das, A. VoicePM: A Robust Privacy Measurement on Voice Anonymity. In Proceedings of the ACM Conference on Security and Privacy in Wireless and Mobile Networks (WiSec), Guildford, UK, 29 May–1 June 2023; pp. 210–221.
53. Ho, S.; Qu, Y.; Gu, B.; Gao, L.; Li, J.; Xiang, Y. DP-GAN: Differentially Private Consecutive Data Publishing Using Generative Adversarial Nets. *J. Netw. Comput. Appl.* **2021**, *185*, 103066. [CrossRef]
54. Torkzadehmahani, R.; Kairouz, P.; Paten, B. Dp-cgan: Differentially Private Synthetic Data and Label Generation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, CA, USA, 16–20 June 2019; pp. 98–104.
55. Jordon, J.; Yoon, J.; Van Der Schaar, M. PATE-GAN: Generating Synthetic Data with Differential Privacy Guarantees. In Proceedings of the 7th International Conference on Learning Representations (ICLR), New Orleans, LA, USA, 6–9 May 2019.
56. Torfi, A.; Fox, E.A.; Reddy, C.K. Differentially Private Synthetic Medical Data Generation Using Convolutional GANs. *Inf. Sci.* **2022**, *586*, 485–500. [CrossRef]
57. Kaltakis, I.; Polyzi, E. Privacy-Preserving Solutions in Blockchain-Enabled Internet of Vehicles. *Appl. Sci.* **2021**, *11*, 9792. [CrossRef]
58. Zhou, W.; Zhang, D.; Han, G.; Zhu, W.; Wang, X. A Blockchain-Based Privacy-Preserving and Fair Data Transaction Model in IoT. *Appl. Sci.* **2023**, *13*, 12389. [CrossRef]
59. Yuan, S.; Yang, W.; Tian, X.; Tang, W. A Blockchain-Based Privacy Preserving Intellectual Property Authentication Method. *Symmetry* **2024**, *16*, 622. [CrossRef]
60. Rathod, T.; Jadav, N.K.; Tanwar, S.; Polkowski, Z.; Yamsani, N.; Sharma, R.; Alqahtani, F.; Gafar, A. AI and Blockchain-Based Secure Data Dissemination Architecture for IoT-Enabled Critical Infrastructure. *Sensors* **2023**, *23*, 8928. [CrossRef] [PubMed]
61. Li, K.; Lohachab, A.; Dumontier, M.; Urovi, V. Privacy Preservation in Blockchain-Based Healthcare Data Sharing: A Systematic Review. *Peer-to-Peer Netw. Appl.* **2025**, *18*, 44–61. [CrossRef] [PubMed]
62. Wu, Y.; Yu, N.; Li, Z.; Backes, M.; Zhang, Y. Membership Inference Attacks Against Text-to-Image Generation Models. In Proceedings of the 11th International Conference on Learning Representations (ICLR), Kigali, Rwanda, 1–5 May 2023.
63. Lian, P.; Cai, Y.; Li, S.; Bao, B. Noise as a Probe: Membership Inference Attacks on Diffusion Models Leveraging Initial Noise. In Proceedings of the 35th USENIX Security Symposium (USENIX Security 26), Baltimore, MD, USA, 12–14 August 2026.
64. Taub, A.K.; Grabovski, F.M.; Amit, G.; Mirsky, Y. LeakBoost: Perceptual-Loss-Based Membership Inference Attack. *arXiv* **2026**, arXiv:2602.05748.
65. Guan, V.; Guépin, F.; Cretu, A.-M.; de Montjoye, Y.-A. A Zero Auxiliary Knowledge Membership Inference Attack on Aggregate Location Data. *arXiv* **2024**, arXiv:2406.18671.
66. Mistareeh, H.; Tennyson, M.; Salameh, B. Enhancing privacy-preserving in vehicular cloud through an incentive-based strategy. *Int. J. Electr. Comput. Eng.* **2024**, *14*, 4078. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.