Privacy-by-Architecture for Passenger Counting in Growing Economies

1st Nthenya Kyatha
Manning College of Info. & Computer Sciences
University of Massachusetts Amherst
Amherst, USA
mkyatha@umass.edu

2nd Nduku Kyatha
Kenya School of Law
Karen, Kenya
kyathanduku@gmail.com

3rd Jay Taneja
Manning College of Info. & Computer Sciences
University of Massachusetts Amherst
Amherst, USA
jtaneja@umass.edu

***Abstract*—Vision-based automatic passenger counting (APC) increasingly relies on biometric-adjacent signals (e.g., ReID embeddings and pose/appearance cues) to remain accurate under crowded, occluded, and poorly lit bus-door conditions. These same signals can enable persistent tracking, creating tension with rapidly evolving data protection regimes in sub-Saharan Africa (SSA) and the broader Global South, especially as electric-bus deployments expand onboard CCTV. We argue that the passenger-counting community underestimates this ReID–anonymisation paradox; accuracy improvements often expand the privacy attack surface. Drawing on deployments in two East African cities, we highlight three under-examined risks: embedding persistence, failure-mode data fabrication, and trajectory inference from count metadata and propose a three-tier Privacy-by-Architecture framework: (1) ephemeral edge processing with mandatory expiry, (2) count-only output contracts, and (3) regulatory alignment with emerging DPAs. We show that these architectural constraints need not reduce counting accuracy, and call for standardised privacy benchmarks beyond face-centric anonymisation.**

***Index Terms*—privacy-preserving analytics, passenger counting, person re-identification, edge computing, data protection, sub-Saharan Africa, soft biometrics**

I. Introduction and Position Statement

Modern vision-based automatic passenger counting (APC) systems are *de facto* soft biometric systems. They extract and process person re-identification (ReID) embeddings, pose keypoints, and appearance descriptors that encode identity-bearing information for each individual boarding or exiting a bus. Despite this, these systems frequently do not receive the same level of privacy scrutiny as facial recognition or gait analysis technologies. The resulting lack of oversight introduces significant risks in growing economies across sub-Saharan Africa (SSA) and the Global South, where APC systems are being deployed at scale on public transit networks amid developing data protection frameworks and limited enforcement capacity.

Position. Privacy should be treated as a primary architectural constraint in APC system design, rather than addressed through post-hoc anonymisation. Growing economies in SSA are particularly vulnerable to privacy failures in transit analytics, yet are also well-positioned to develop privacy-by-architecture approaches that operate under resource constraints. The proposed solutions are intended to be transferable to other growing economies with comparable regulatory and deployment environments.

Although architectural safeguards can mitigate privacy risks, they do not substitute for legal compliance obligations, including establishing a lawful basis, providing notice, and upholding data subject rights. We present Privacy-by-Architecture as a way to make compliance feasible under resource and enforcement constraints, not as an alternative to it.

A. The Growing-Economy Context

AI-driven surveillance is expanding across the SSA, often through foreign vendor systems with limited local oversight [7]. A 2020 brief reported that, as of June 2020, at least 12 African countries were using Huawei surveillance technology through “Safe City” deployments [14]; later analyses suggest deployments in 16+ countries, with totals varying by definitions and sources [15]. Meanwhile, although 44 of 55 African nations have enacted data protection laws and at least 38 have established DPAs, enforcement capacity is uneven: 6 countries have laws without operational regulators, and 11 have yet to legislate [13].

Recent enforcement actions underscore the significance of robust data protection. In May 2025, Kenya’s High Court mandated the permanent deletion of iris-scan data collected without a Data Protection Impact Assessment (DPIA) and informed consent, establishing that the absence of a DPIA can independently invalidate biometric processing [8].

Concurrently, the transition from diesel to electric vehicles is accelerating the deployment of onboard video infrastructure. Kenya has announced plans to fully electrify public transport by 2027 [6], and by August 2025, over 100 locally assembled electric buses had been delivered in Kenya and Rwanda [3]. Many of these platforms are equipped with factory-installed closed-circuit television (CCTV) systems by default; however, policies regarding data retention and cross-border access remain ambiguous. This infrastructure amplifies privacy risks alongside modern automatic passenger counting (APC) pipelines that utilise YOLO-based detection [11], tracking, and ReID embeddings to maintain accuracy in unconstrained environments.

B. Deployment-Specific Challenges

The deployment context in SSA shapes the privacy-utility trade-off differently than in European or North American settings. Typical conditions involve dashcam-quality cameras with wide-angle lenses, severe backlighting at bus doors, intermittent cellular connectivity, and edge devices with limited computational capacity. These constraints paradoxically increase reliance on appearance-based features. In the absence of face-level resolution, the system must use full-body ReID embeddings, clothing descriptors, and pose cues, all of which are more difficult to anonymise than facial features alone.

C. Novelty and Contribution

This work addresses a previously overlooked gap at the intersection of APC and soft-biometric tracking. While modern APC systems employ ReID embeddings to maintain count accuracy under occlusion, the literature generally treats outputs as aggregate counts and does not subject them to surveillance scrutiny. The ReID–anonymisation paradox is formalised: identity-bearing representations that prevent double-counting simultaneously enable re-identification. Based on deployments in two East African cities, the contributions include: (i) a deployment-grounded threat model and taxonomy of APC-specific privacy risks (embedding persistence, failure-mode fabrication, and metadata-only trajectory inference); (ii) a Privacy-by-Architecture framework that enforces time-bounded representations on-device and a count-only output contract at the system boundary; and (iii) a concrete path toward standardisation through an APC Privacy Benchmark that measures leakage from embeddings and count/metadata streams. Collectively, these contributions shift APC privacy from post-hoc anonymisation to enforceable architectural constraints.

II. The ReID–Anonymization Paradox

ReID embeddings, which underpin modern multi-object tracking in APC systems, encode enough appearance information to serve as soft biometric templates. The discriminative properties that enable these embeddings to resolve identity switches and recover tracks after occlusions also facilitate cross-session and cross-camera re-identification. We term this phenomenon the *ReID–anonymisation paradox*: accurate passenger counting necessitates identity-bearing features, yet retaining them introduces surveillance risks.

A. What APC Systems Actually Capture

In practice, a modern APC pipeline on bus-mounted cameras produces identity-bearing intermediates in addition to the final count. It computes person ReID embeddings (often 128–2048D appearance vectors) that encode body shape, clothing texture, and carried objects and can function as implicit biometric templates [16]. Many systems also estimate pose keypoints (or short pose sequences) that preserve body proportions and gait-adjacent cues. For counting and occlusion recovery, the pipeline maintains spatiotemporal tracklets linked across frames (and sometimes across cameras), along with metadata such as timestamps, camera IDs, ROI/door-crossing events, and confidence scores. In operational settings, further artefacts are often retained for association and debugging, including bounding boxes, cropped person images, track IDs, ReID similarity scores, and short-lived buffers or logs. When fused with vehicle context such as GPS-derived stop IDs or door-state signals, these time–location traces can support reconstruction of individual travel patterns even when raw video is not exported. In SSA transit, distinctive items (market goods, uniforms, work equipment) increase appearance discriminability, enabling re-identification without facial detail.

B. Why Naive Anonymisation Fails

Face blurring, a commonly discussed anonymisation technique, is insufficient for APC because ReID relies on full-body appearance cues rather than just facial pixels. Prior work shows that even when RGB facial detail is suppressed (e.g., event-based anonymisation), body-level signals can still support re-identification [1]. More broadly, edge-friendly anonymisation methods face persistent tradeoffs between privacy and performance for downstream video analytics tasks [2].

Skeleton-only processing is also inadequate in crowded doorways: removing appearance cues weakens occlusion recovery and short-term re-association. Finally, many anonymisation benchmarks are face-centric and built on datasets from controlled settings, whereas bus CCTV in SSA is low-resolution and unconstrained, which increases reliance on holistic clothing/silhouette cues that are harder to anonymise.

C. Evidence from Deployment

Field deployments of dual-camera passenger-counting systems on public transit buses demonstrate that tracking accuracy depends significantly on features with re-identification potential. In these deployments, extreme backlighting at bus doors frequently eliminates face-level cues, necessitating reliance on full-body appearance signals such as silhouette, clothing, and carried items. This observation is consistent with prior video-based APC studies, which emphasize that bus CCTV is captured under uncontrolled conditions that challenge detection and tracking pipelines [17], [19]. These appearance descriptors are exploited by modern ReID systems for cross-camera matching [16]. During crowded boarding events, appearance matching across fragmented detections is essential, as a single individual may be split into multiple tracks that require re-association using ReID similarity [18]. Such scenarios represent the default operating conditions for bus-mounted APC systems in growing economies.

III. Under-Examined Privacy Risks

While APC systems are primarily designed for aggregate counting, the architecture of modern tracking pipelines introduces three distinct categories of privacy risk. These risks are particularly pronounced in emerging economies, where device security tends to be weaker, data governance frameworks are less mature, and documented cases of misuse by state or non-state actors are prevalent [7].

A. Embedding Persistence Risk

ReID embeddings stored in tracking buffers, ID repair queues, and re-association caches may persist well beyond the period required for counting. In multi-camera configurations, such as those with both entrance and exit cameras, embeddings must remain accessible across camera views to enable cross-camera biometric matching. If an edge device is compromised or if embedded buffers are logged for debugging, these caches may inadvertently form a biometric database.

This risk is especially pronounced in SSA, where edge devices installed on buses often have limited physical security, are accessible to third-party maintenance personnel, receive infrequent firmware updates, and are procured through complex supply chains with multiple intermediaries. Collectively, these factors expand the attack surface for embedding extraction.

B. Failure-Mode Data Fabrication

Tracking failures introduce privacy risks that are frequently overlooked. *Ghost injection*, the creation of false-positive identities from background clutter or transient detections, results in biometric records for individuals who do not exist. *ID fragmentation* occurs when a single individual is split into multiple tracked identities, each with stored appearance features, thereby increasing the biometric footprint beyond the actual number of individuals present.

These failure modes have significant privacy implications. In environments where biometric data has been repurposed for political surveillance [7], even fabricated records pose risks. Artificial biometric traces linked to specific locations and timestamps may be erroneously attributed to actual individuals.

C. Trajectory Inference from Metadata

Even if ReID embeddings are removed, APC outputs (stop-level boarding/alighting counts with timestamps) can enable trajectory inference when linked with auxiliary datasets such as fare-payment logs, mobile network traces, or other CCTV streams. In many SSA settings, fares are increasingly paid via mobile money platforms that record persistent payer identifiers (e.g., phone numbers) linked to verified subscriber identities through Know Your Customer (KYC) processes. Prior work shows that trajectories alone can be sufficient to re-identify users at scale [10]. The risk is amplified in informal transit networks with fewer alternative routes, which reduces the combinatorial space and makes per-stop count sequences more attributable than in dense metro systems.

D. Factory-Fitted Cameras and Third-Party Data Flows

These factory-fitted cameras raise distinct concerns beyond aftermarket APC systems, where the deployer controls the entire data pipeline. First, the degree of OEM access (from manufacturers such as BYD and Higer [4]) to video feeds or derived analytics is often ambiguous, increasing the likelihood of cross-border data transfers to jurisdictions with inadequate data protection. Second, bus operators may misuse camera feeds for passenger profiling, sell behavioural data to advertisers, or engage in informal surveillance, all of which exceed the intended safety function. Third, factory-fitted cameras typically record the entire bus interior continuously, rather than focusing solely on the door region during boarding, resulting in more comprehensive biometric datasets than those generated by purpose-built APC systems. As SSA countries plan to deploy thousands of electric buses by 2027, this surveillance infrastructure is poised to outpace regulatory oversight [3], [6].

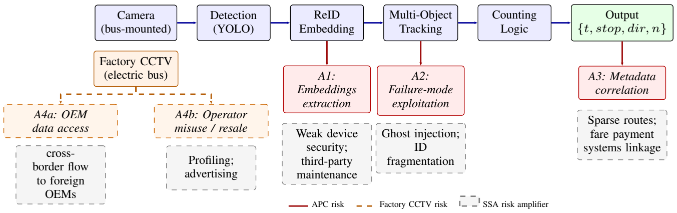

Fig. 1. Threat model for vision-based transit surveillance. Attack vectors A1–A3 (solid, red) target purpose-built APC pipelines at the embedding, tracking, and output stages. Vector A4 (dashed, orange) targets factory-fitted CCTV on electric buses, where data may flow to foreign OEMs or be repurposed by operators. Gray dashed boxes indicate growing-economy risk amplifiers.

IV. Proposed Framework: Privacy-by-Architecture

Privacy must be integrated into system architecture from sensor to output, rather than addressed as an add-on through anonymisation filters. This integration is especially critical in growing economies, where resource constraints often preclude post-hoc compliance auditing. The proposed three-tier *Privacy-by-Architecture* framework (see Figure 2) is designed for deployment within the computational, connectivity, and institutional constraints characteristic of SSA transit systems.

A. Tier 1: Ephemeral Edge Processing

Principle. Biometric-adjacent representations (e.g., ReID embeddings, pose sequences) *must not* leave the edge device under the system's operating contract, and must be time-bounded.

Mechanism. A hard time-to-live (TTL) is enforced on ReID embedding caches and any tracking/association buffers *from track initiation*. Once a track is finalised (i.e., a boarding/alighting decision is made), embeddings are destroyed rather than retained for the remainder of a route. When full secure wiping is infeasible on heterogeneous hardware, TTL can be implemented via *cryptographic erasure* (e.g., encrypting embeddings with an ephemeral key and destroying the key on expiry). Diagnostic logs are subject to the same TTL constraints and must not, by default, contain raw embeddings.

Tradeoff. Ephemeral processing restricts post-hoc accuracy audits and detailed remote debugging. This limitation is partially addressed by retaining only aggregate statistics (and optional, coarse, anonymised heatmaps) for validation.

Growing-economy advantage. Edge-first processing is frequently the default in SSA transit deployments because intermittent connectivity reduces the feasibility of continuous upstream data transmission. However, intermittent connectivity alone does not ensure privacy, as store-and-forward uploads and debug logs may still occur. The proposed framework formalises edge-only constraints and TTL deletion as enforceable privacy guarantees.

B. Tier 2: Count-Only Output Contracts

Principle. The system's external interface is limited to aggregate count reports; rich per-person outputs are contractually and technically disallowed.

Mechanism. API-level enforcement restricts transmitted data to the minimal tuple: {`timestamp, stop_id, direction, count`}. No embeddings, bounding boxes, track identifiers, or individual trajectories are exported.

Scope note. In low-ridership settings, even stop-level counts can become linkable when combined with auxiliary datasets. To mitigate low-n leakage, release counts only after aggregation thresholds are met and, where needed, add calibrated noise following differential privacy methods [5].

Tradeoff. Count-only contracts reduce remote diagnostics, mitigated by encrypted, access-controlled on-device logs with time-limited retention aligned with Tier 1 (and excluding raw embeddings by default).

Factory-fitted camera gap. Factory-integrated CCTV on electric buses is often managed for safety and maintenance outside the APC pipeline, so it may retain or expose richer video streams and bypass count-only and short-retention controls. We recommend extending Tier 2 procurement requirements to all onboard cameras, requiring explicit retention limits, role-based access controls with audit logging, and restrictions on secondary use and cross-border transfers.

C. Tier 3: Regulatory Alignment

Principle. Map each technical control to specific legal obligations under data protection frameworks so that architectural privacy translates into compliance.

Scope note. Ephemerality reduces the surface area for rights requests on stored biometric-adjacent templates, but increases the importance of transparency about what is retained, for how long, and for what purpose. Architectural safeguards reduce exposure, but do not substitute for lawful basis, notice, and data-subject rights obligations.

The framework is mapped to two representative data protection regimes in East Africa, both inspired by the General Data Protection Regulation (GDPR). This mapping supports portability to the increasing number of African countries with enacted data protection laws and to GDPR-aligned frameworks in other regions.

The landmark ruling in *Katiba Institute v. Worldcoin Foundation* [8] reinforces the urgency of Tier 3: the court found that biometric data collection without a DPIA was unlawful *regardless of the collector's stated purpose* and ordered permanent data deletion under DPA supervision. This precedent applies directly to APC operators who process ReID embeddings without formal privacy assessments. We recommend that all transit authorities deploying vision-based APCs register as data controllers and conduct DPIAs addressing ReID embedding processing.

TABLE I
Regulatory compliance mapping for the framework.DPA-K: Data Protection Act, Kenya (2019); DPA-R: Law N^{O}058/2021, Rwanda.

Framework Tier	Statutory hooks	How the tier helps satisfy the obligation

Tier 1: Ephemeral Edge	**DPA-K:**
	Sec. 25(d),(g);
	Sec. 39(1)–(2);
	Sec. 41(1)–(4);
	Sec. 42
	DPA-R:
	Art. 37(5);
	Art. 38(1);
	Art. 47
		Minimisation & storage limitation: keep identity-bearing signals only as long as necessary for the counting purpose.
		Deletion at expiry: delete, erase, anonymise/pseudonymise once no longer necessary.
		Security by design: implement technical/organisational measures (e.g.encryption/pseudonymisation) proportionate to risk.

Tier 2: Count-Only Output	**DPA-K:**
	Sec. 25(b)–(d),(h);
	Sec. 41(3);
	Sec. 48;
	Sec. 49(1)
	DPA-R:
	Art. 37;
	Art. 49
		Purpose limitation: restrict outputs to what is necessary for the stated purpose.
		Data minimisation by default: expose only the minimal schema needed.
		Cross-border (conditional): transferring personal (or sensitive) data outside Kenya must satisfy statutory transfer conditions/safeguards (Sec. 25(h); Sec. 48; and for sensitive data, Sec. 49(1)).

Tier 3: (DPIA & breach governance)	**DPA-K:**
	Sec. 31;
	Sec. 43;
	Sec. 24(1)(b)
	DPA-R:
	Art. 38(3);
	Art. 40–41;
	Art. 43;
	Art. 53
		High-risk processing governance: Conduct DPIA prior to processing and document safeguards/controls.
		Breach response: notify the Commissioner within statutory timelines where unauthorised access/acquisition creates a real risk of harm.
		Accountability: designate a DPO where processing involves regular and systematic monitoring.

Note: The Act defines *personal data* to include information enabling identification directly or indirectly (including via location/time linkage); aggregated outputs may still be personal data when linkable to identifiable individuals [9], [12].

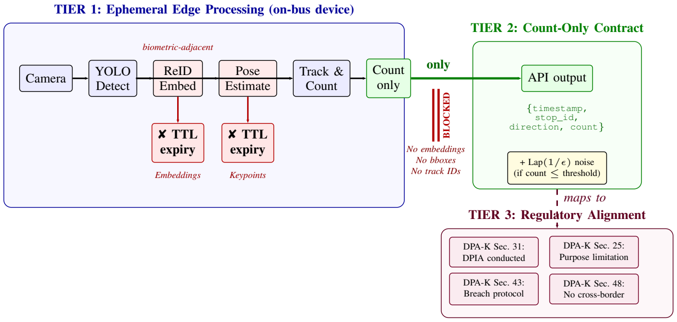
TIER 1: Ephemeral Edge Processing (on-bus device)
biometric-adjacent
Camera
YOLO Detect
ReID Embed
Pose Estimate
Track & Count
Count only
✘ TTL expiry
✘ TTL expiry
Embeddings
Keypoints
only
BLOCKED
No embeddings
No bboxes
No track IDs
TIER 2: Count-Only Contract
API output
{timestamp, stop_id, direction, count}
+ Lap(1/ϵ) noise (if count ≤ threshold)
maps to
TIER 3: Regulatory Alignment
DPA-K Sec. 31: DPIA conducted
DPA-K Sec. 25: Purpose limitation
DPA-K Sec. 43: Breach protocol
DPA-K Sec. 48: No cross-border

Fig. 2. The Privacy-by-Architecture framework. Biometric-adjacent data (red-tinted blocks) exists only within Tier 1 on the edge device and is destroyed via TTL expiry (✘). Only the aggregate count tuple crosses the Tier 2 boundary. Tier 3 maps each technical control to Kenya DPA provisions.

D. Scope note (technical controls vs. compliance)

The proposed tiers are technical and procurement constraints that reduce exposure to biometric-adjacent signals; they are not a substitute for compliance. Even with strict TTL deletion and count-only outputs, operators must still identify a lawful basis, provide notice, and meet obligations on retention, data subject rights, and accountability under the applicable DPA. We position Privacy-by-Architecture as a way to make compliance feasible under resource constraints, not as an alternative to it.

V. Preliminary Feasibility Analysis

The feasibility of the proposed framework is assessed through an architectural analysis of an actual APC system implemented on public transit buses in two East African cities. This analysis evaluates whether each tier introduces fundamental barriers to counting accuracy.

A. Tier 1: Is Ephemeral Processing Compatible with Counting?

A central question is whether ReID embeddings must persist beyond a single boarding or alighting event. In a typical APC pipeline, embeddings associate detections of the same individual across consecutive frames to prevent double-counting. Once an individual has been tracked from the door threshold to a seated or standing position, the counting decision is finalised and the embedding is no longer required.

In practice, the required time window is determined by dwell time (the duration a bus spends at a stop with doors open). In informal transit systems across Sub-Saharan Africa, dwell times are generally short because operators minimise stop time to maintain competitive positioning. A TTL of several tens of seconds would accommodate most boarding events. Extended stops caused by traffic or terminal layovers are edge cases in which reverting to detection-based counting without ReID results in graceful degradation rather than system failure. Formal measurement of dwell-time distributions in specific deployment contexts would substantiate this argument and represent a direction for future research.

B. Tier 2: Does Count-Only Output Sacrifice Utility?

In a well-designed APC system, the counting decision is based on an individual crossing a directional threshold, such as moving from outside to inside the door region. The final output, which is a count increment, requires only confirmation that a crossing occurred, without reference to the individual's identity, appearance, or trajectory. Consequently, the count-only output contract at Tier 2 (`timestamp, stop_id, direction, count`) preserves all information necessary for transit planning, including route-level ridership, stop-level demand, and temporal patterns, while eliminating any data that could identify individuals.

The primary drawback of a count-only output is reduced debuggability. When counting errors occur, operators cannot remotely replay per-track data to diagnose the underlying cause. However, this trade-off is considered acceptable and can be mitigated through on-device diagnostic logging, with encrypted, time-limited access provided to authorised maintenance personnel.

C. Tier 3: Compliance Gap Analysis

An audit of the deployed system against the Kenya DPA provisions, as mapped in Table I, identified three compliance gaps: (a) no formal TTL was enforced on embedding buffers, resulting in embeddings persisting for the duration of a route segment rather than being deleted after track finalization; (b) diagnostic logs contained track-level data, including bounding box coordinates and ReID similarity scores; and no DPIA had been conducted for the ReID processing component. Gaps (a) and (b) are engineering issues that require moderate implementation effort, while gap necessitates institutional action by the transit authority. None of these gaps constitutes a fundamental architectural barrier; rather, they reflect the absence of privacy as a design-time constraint, which this paper aims to address.

VI. Discussion and Call to Action

A. Addressing Counter-Arguments

"Passenger counting is not biometric surveillance.‘' The determining factor for regulatory scope is the *capability* for re-identification, rather than the *intent*. Data protection laws in Kenya and Rwanda define personal data broadly as any information relating to an identified or *identifiable* natural person [9], [12]. ReID embeddings designed to distinguish between individuals clearly fall within this definition, regardless of the system's stated purpose. The Worldcoin precedent [8] demonstrates that courts assess biometric processing based on its technical capabilities rather than its declared objectives.

"Edge processing already protects privacy.‘' Edge processing alone is insufficient without formal guarantees. Device seizure, firmware exploits, or insider access can expose buffered embeddings. In SSA contexts, where physical device security is often weaker, and maintenance is conducted by third parties, this counter-argument is particularly vulnerable.

"Privacy compliance will slow APC deployment in regions that need it most.‘' In fact, the opposite is true. The Worldcoin suspension demonstrated that *failing* to address privacy proactively can halt deployments entirely. Proactive compliance fosters public trust and regulatory goodwill, thereby accelerating rather than hindering adoption. Establishing privacy infrastructure during the initial deployment of APC systems is significantly more cost-effective than retrofitting after a compliance crisis.

"Growing economies have bigger problems than transit data privacy.‘' This perspective establishes a misleading hierarchy of priorities. As Kenya pursues full electrification of public transport, the window to integrate privacy into transit infrastructure prior to large-scale deployment is closing. The European experience of retroactively applying GDPR to existing surveillance infrastructure provides a costly cautionary example. Integrating privacy protections from the beginning is not a luxury but a more cost-effective strategy.

B. Limitations

The proposed framework remains preliminary and requires validation through threat modelling, including adversarial red-teaming of edge devices and data flows. The empirical evidence is based on informal transit deployments in two East African cities and may not generalise to bus rapid transit, metro, paratransit operations, or fleets with different camera placements, OEM CCTV systems, or connectivity configurations. We have not evaluated all sensing modalities, such as infrared, thermal imagery, or audio from onboard microphones. The legal mapping is limited to current statutory text; regulatory guidance, enforcement capacity, and case-law precedent in SSA remain inconsistent and subject to ongoing change. Key design parameters, including time-to-live, aggregation thresholds, and differential privacy noise, have not been systematically tuned through utility–privacy evaluation.

C. Call to Action

All stakeholders in the transit video analytics pipeline in SSA, including transit agencies and the research community, are encouraged to treat privacy as a first-class design and procurement requirement rather than a post-hoc add-on. Stakeholders should develop deployment-realistic privacy benchmarks that quantify soft-biometric leakage from ReID embeddings, pose, and trajectory metadata under low-resolution, uncontrolled bus CCTV conditions; standardise embedding TTL and secure destruction protocols for edge-deployed tracking (with reference implementations for ARM-class hardware common in SSA); and publish open-source DPIA templates tailored to vision-based transit analytics under emerging data protection regimes. Funders are advised to prioritise interdisciplinary work that connects computer vision, data protection law, and transit planning to ensure that technical controls align with operational and legal obligations. Procurement authorities should require privacy-aware contracts for electric bus fleets, ensuring that factory-fitted CCTV systems include explicit governance terms such as retention limits, role-based access controls, audit logging, and restrictions on secondary use and cross-border access.

VII. Conclusion

Vision-based passenger counting systems in growing economies occupy a largely unexplored intersection of biometric technology, surveillance capabilities, and evolving data protection law. The concurrent adoption of electric buses with factory-installed CCTV further intensifies these challenges by embedding surveillance infrastructure throughout transit fleets. This work identifies the ReID–anonymisation paradox as the central conflict between counting accuracy and passenger privacy, and introduces a three-tier Privacy-by-Architecture framework to address this issue. The framework comprises ephemeral edge processing, count-only output contracts, and regulatory alignment. The feasibility analysis shows that none of the three tiers imposes fundamental limitations on counting accuracy. Embeddings are necessary only for the duration of a boarding event, and the counting process does not require retention of per-person data. The rapidly developing economies of SSA, where automated passenger counting systems and CCTV-equipped electric buses are being widely deployed alongside new data protection regulations, present both the most urgent need and the most appropriate environment for evaluating these approaches.

References

[1] A. Ahmad, P. Morerio, and A. Del Bue, “Person Re-identification without Identification via Event Anonymization,” in *Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV)*, 2023, pp. 11132–11141.

[2] M. W. Asres, L. Jiao and C. Walter Omlin, “Low-Latency Video Anonymization for Crowd Anomaly Detection: Privacy Versus Performance,” in IEEE Transactions on Information Forensics and Security, vol. 21, pp. 1-16, 2026, doi: 10.1109/TIFS.2025.3630347.

[3] “BasiGo delivered 100 electric buses in Kenya and Rwanda,” *electrive.com*, Aug. 7, 2025. [Online]. Available: https://www.electrive.com/2025/08/07/basigo-delivered-100-electric-buses-in-kenya-and-rwanda/ [Accessed: Feb. 10, 2026].

[4] “Chinese EV tech in Kenya’s mass transport: The BasiGo story,” *The China-Global South Project*, May 30, 2024. [Online]. Available: https://chinaglobalsouth.com/2024/05/30/chinese-ev-tech-in-kenyas-mass-transport-the-basigo-story/ [Accessed: Feb. 10, 2026].

[5] C. Dwork, F. McSherry, K. Nissim, and A. Smith, “Calibrating noise to sensitivity in private data analysis,” in *Proc. Theory of Cryptography Conf. (TCC)*, 2006, pp. 265–284.

[6] “Kenya plans to fully electrify its bus fleet by 2027,” *electrive.com*, Apr. 17 2024. [Online]. Available: https://www.electrive.com/2024/04/17/kenya-plans-to-fully-electrify-its-bus-fleet-by-2027/ [Accessed: Feb. 10, 2026].

[7] A. Gwagwa, E. Kraemer-Mbula, N. Rizk, I. Rutenberg, and J. De Beer, “Artificial intelligence (AI) deployments in Africa: Benefits, challenges and policy dimensions,” *The African Journal of Information and Communication (AJIC)*, vol. 26, pp. 1–28, 2020.

[8] *Republic v. Tools for Humanity Corporation (US) & 8 others; Katiba Institute & 4 others (Ex parte Applicants); Data Privacy & Governance Society of Kenya (Interested Party)*, Judicial Review Application No. E119 of 2023, [2025] KEHC 5629 (KLR) 5 May 2025 (Judgment), High Court at Nairobi (Milimani Law Courts). [Online]. Available: https://new.kenyalaw.org/akn/ke/judgment/kehc/2025/5629/eng%402025-05-05

[9] Republic of Kenya, *The Data Protection Act, No. 24 of 2019*, Kenya Gazette Supplement No. 181, 2019. Available: https://kenyalaw.org/kl/fileadmin/pdfdownloads/Acts/2019/TheDataProtectionAct__No24of2019.pdf

[10] A. K. Mishra, M. Cunche, and H. H. Arcolezi, “Breaking anonymity at scale: Re-identifying the trajectories of 100K real users in Japan,” *arXiv preprint* arXiv:2506.05611, 2025.

[11] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779–788.

[12] Republic of Rwanda, *Law No 058/2021 of 13/10/2021 Relating to the Protection of Personal Data and Privacy*, October 15, 2021. Available: https://www.risa.gov.rw/data-protection-and-privacy-law

[13] Digital Policy Alert, “The Year of the Teeth: Data Protection in Africa Roundup, 2025, Projections for 2026,” Jan. 11, 2026. [Online]. Available: https://digitalpolicyalert.org/blog/data-protection-in-africa-roundup. [Accessed: Feb. 10, 2026].

[14] B. Jili, “Chinese Surveillance Tools in Africa,” Research Brief No. 8/2019, Jun. 30, 2020. [Online]. Available: https://cld.web.ox.ac.uk/file/678231. [Accessed: Feb. 10, 2026].

[15] C. C. Passalacqua and C. Polito, “The Chinese Supply of Surveillance Technology to Africa: Going Beyond the Authoritarian Bias,” Luiss School of Government Working Paper Series, SOG-WP7/2024, Feb. 2024. [Online]. Available: https://sog.luiss.it/sites/sog.luiss.it/files/The%20Chinese%20Supply%20of%20Surveillance%20Technology%20to%20Africa.pdf. [Accessed: Feb. 10, 2026].

[16] M. Ye, J. Shen, G. Lin, T. Xiang, L. Shao, and S. C. H. Hoi, “Deep learning for person re-identification: A survey and outlook,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 6, pp. 2872–2893, Jun.2022, doi: 10.1109/TPAMI.2021.3054775.

[17] Y.-W. Hsu, T.-Y. Wang, and J.-W. Perng, “Passenger flow counting in buses based on deep learning using surveillance video,” *Optik*, vol. 202, Art. 163675, 2020, doi: 10.1016/j.ijleo.2019.163675.

[18] C. Labit-Bonis, J. Thomas, and F. Lerasle, “Visual and automatic bus passenger counting based on a deep tracking-by-detection system,” HAL preprint hal-03363502, 2021.

[19] C. McCarthy, H. Ghaderi, F. Martí, P. Jayaraman, and H. Dia, “Video-based automatic people counting for public transport: On-bus versus off-bus deployment,” *Computers in Industry*, vol. 164, Art. 104195, 2025, doi: 10.1016/j.compind.2024.104195.