Letter from the Editor-in-Chief

Privacy in AI is having its "everything, everywhere" moment. It shows up in procurement checklists, product reviews, incident postmortems, and policy debates, and it is often discussed with the urgency of an existential risk and the practicality of a compliance chore. That contradiction is not a sign of confusion so much as it is a sign that we are living through a transition. The tools and institutions that once made privacy feel manageable were built for a world in which data use looked like a bounded transaction: a service collected information, stored it, and used it for an identifiable purpose. In the AI era, data does not merely support a service; it becomes part of a capability that can be reused, repurposed, and redeployed, long after the original context has faded.

That is why it is becoming harder to believe in privacy as a moment—a notice, a click, a checkbox, a clause. When people are asked to make meaningful decisions through low-signal interfaces, under time pressure, in environments engineered for throughput rather than deliberation, "choice" starts to resemble a ritual rather than an exercise of agency. Even when consent is present, it rarely carries predictive power about what will happen next, because modern AI systems blur the lines between primary use and secondary use. Data flows into pipelines; pipelines feed models; models are integrated into products; products generate logs; logs become training material; and vendors, tools, and internal services become links in a chain that few individuals can see, much less control.

The papers in this issue suggest a sturdier thesis: privacy in AI will be won or lost by infrastructure. By "infrastructure," I mean defaults that do not require constant vigilance from the individual, controls that remain meaningful as systems scale and get outsourced, and accountability that follows data and models across their lifecycle rather than concentrating at the moment of collection. The old world asked whether privacy disclosures were complete and whether consent was obtained. The new world has to ask whether the system is constructed so that restraint is the normal path, misuse is harder than proper use, and the consequences of failure are detectable before they become irreversible.

One way to see the limits of the older framing is to notice how often privacy debates get stuck on a narrow question: did the model memorize sensitive training examples, and can that information be extracted? That question matters, but it is no longer the whole story. The paper *Beyond Data Privacy: New Privacy Risks for Large Language Models* by Du et al, offers a useful reframing by widening the threat model. In practice, privacy risk emerges not only from what was in the training set, but from the way models are deployed, connected, and operated. Prompts and outputs may be logged. Retrieval systems may introduce sensitive context. Tool integrations and plugins may route data to third parties. An application can be carefully designed in one layer and quietly undermine privacy in another. The most unsettling element of this expanded frame is that privacy harm can arise even when nothing is "leaked" in the classic sense, because models can be used to infer, profile, and amplify invasive behavior. In other words, the privacy story shifts from a single vulnerability to an exposure surface spanning the full stack.

Once you accept that, another gap in conventional thinking becomes obvious. We often talk about privacy as if it were mainly a question of who can see the data. Yet modern AI systems increasingly depend on outsourced infrastructure, external platforms, and complex vendor ecosystems, and that raises a second question that is just as fundamental: can we trust what computation happened, where, and under what constraints? The paper *Data Privacy and Computation Integrity in Machine Learning Scenarios* insists that privacy promises become brittle when integrity is treated as an afterthought. If training, storage, or evaluation is delegated to an environment you do not fully control, the question is not merely whether the data was protected in transit or at rest, but whether the computation was performed correctly, completely, and in a way that can be audited. In real deployments, privacy failures are often caused less by cinematic adversaries and more by mundane realities: misconfiguration, partial failures, permissive logging, quiet pipeline drift, and vendor-side changes that are hard to observe from

the outside. Integrity mechanisms, verification patterns, and auditable controls do not make systems perfect, but they shift privacy from hope to evidence. Privacy without integrity is a promise that survives only in the best-case scenario.

If infrastructure is the right lens, then privacy by default must also be understood as a systems design problem rather than an algorithmic checkbox. Federated learning, for instance, is frequently introduced as a simple idea: keep data local and ship updates. But the paper *Privacy-Preserving Federated Large Language Models* by Xu et al is valuable precisely because it refuses to sell simplicity. At the scale of large language models, the tensions among privacy, utility, and efficiency are not abstract. If you push too hard on privacy, performance may degrade or training may become unstable under heterogeneous data. If you prioritize utility, you may invite leakage through updates. If you optimize for efficiency, you may weaken the very redundancy and aggregation that make privacy protections workable. The paper reads as an argument for intellectual honesty: privacy engineering is tradeoff engineering. The goal is not to deny tradeoffs, but to make them explicit, measurable, and governable.

That idea – governable tradeoffs – is where "infrastructure" stops being a metaphor. Governance requires enforcement. Enforcement requires permissioning. And permissioning, in large-scale AI services, is inseparable from identity and access control. It is easy to categorize managed identities and authorization systems as "security plumbing" and treat them as adjacent to privacy. In practice, they are privacy. Access drift is a privacy leak. Ambiguous service identity is a privacy risk. Mis-scoped tokens are privacy incidents waiting to happen. The paper *Hyper-Scale Managed Identities and Access Control*, by Alagenchev et al, underscores how much modern systems depend on reliable identity, scalable authorization, and strong assurance about who is calling what. In the AI era, where models become shared services consumed by many clients, privacy cannot be protected if the system cannot express and enforce who may use which model, with what data, under what constraints. The boundary between privacy and security is not disappearing, but it is becoming increasingly artificial at the level where systems actually fail.

Of course, infrastructure is not only about locks; it is also about measurement. What can be observed can be governed, and what cannot be observed becomes a matter of trust. Yet measurement is only as useful as the reality it reflects. Formal guarantees can be technically correct while socially misleading if they are not aligned with how risk manifests in practice. Two contributions in this issue, in different ways, press toward the same lesson: privacy claims must match real units of harm and real deployment contexts. Mahloujifar et al, in *Optimal Group Privacy for DP-SGD*, push beyond the idea that privacy is only about a single record. Many harms accrue at the level of groups, cohorts, households, or communities; user participation itself can be sensitive; and correlated records can amplify exposure. Meanwhile, Mokhtari et al, in *Rethinking Benchmarks for Differentially Private Image Classification*, challenge the tendency to benchmark methods in settings that are convenient rather than representative of privacy-critical domains. Benchmarks shape what we optimize, what we deploy, and what we celebrate. If they fail to reflect the stakes and constraints of real-world settings, we risk building methods that look strong on paper while leaving the highest-risk contexts underserved.

Stepping back, the connective tissue across these papers is not a single mechanism, but a shared insistence that privacy must move upstream and become operational. The path forward is less about asking individuals to shoulder the burden of prediction and vigilance, and more about building systems in which restraint is the default behavior of the pipeline. It is less about treating compliance as a final gate, and more about structuring interdisciplinary ownership so that privacy requirements shape architecture choices early and remain meaningful as systems evolve. It is less about a narrow focus on training data and more about end-to-end thinking that includes deployment realities, integrations, logging, and misuse. It is less about privacy as a promise and more about privacy as an observable property supported by integrity, identity, and measurement.

That framing also clarifies what progress should look like. Progress is a world in which collecting less

is not an act of heroism, but the easiest path. It is a world in which provenance, retention, and purpose are not buried in policy documents, but expressed in system behavior and enforced by controls. It is a world in which outsourcing does not dissolve responsibility, because privacy and integrity constraints survive across vendor boundaries. It is a world in which "privacy-preserving" is not a marketing adjective, but a claim that can be scrutinized with shared benchmarks, realistic assumptions, and guarantees that align with how harm is experienced.

The call to action, then, is not to wait for a single breakthrough. It is to build the missing privacy supply chain. Researchers can prioritize work that is system-aware and deployment-relevant, connecting formal guarantees to operational realities like identity, access control, logging, and outsourcing. Practitioners can stop treating privacy as something to be validated late and instead embed it into data pipelines, model lifecycle management, and continuous monitoring, with the same seriousness we apply to reliability and safety. Institutions can push accountability upstream, encouraging norms that make provenance, retention, and permissioning central rather than peripheral.

If we do that work, privacy in the AI era stops being a rear-guard action. It becomes a discipline of construction: designing systems that can learn and serve at scale without turning personal data into a permanent liability. The papers in this issue do not claim that future has arrived, but they make it easier to see what it would take to build it. That is the real opportunity before us, and it is also the test that will define what comes next.

Haixun Wang
EvenUp