

From the “Real Me” to the “Statistical Me”:

Why Privacy Protection Must Now Govern Data, Representation, and Action

Heng Xu (<https://orcid.org/0000-0001-5642-6543>)

Mendoza College of Business, University of Notre Dame, Notre Dame, IN, USA

Abstract: *This article argues that privacy protection should move beyond protecting the “real me” as the holder of personal data and toward protecting the “statistical me”: the inferred, model-generated version of myself that AI systems construct, remember, and act upon.*

Imagine an organization deploying an artificial intelligence (AI) copilot [i.e., large language model (LLM)] for everyday work. To be useful, it connects to email, calendars, chat, documents, tickets, and meeting transcripts. At first, the privacy questions look familiar. What data can it access? How long is it retained? Who else can see it? Are the logs encrypted? Can employees opt out? These are the questions privacy and security professionals know how to ask.

But these questions no longer reach the center of the system. A useful copilot does more than retrieve information. Around the model sit memory features, retrieval indices over your documents, conversation histories, and stored user profiles. Through those layers, the system accumulates inferential signals about who matters to you in the organization, which tasks you delay, what tone you respond to, when you seem uncertain, and how you handle pressure. These inferences need not live inside the model (much less in its weights) to shape every subsequent interaction. In-context conditioning over whatever the system retrieves and remembers (e.g., documents, profiles, prior turns) is enough (Brown et al., 2020). The system can then use those inferences to personalize outputs, rank priorities, flag anomalies, route decisions, or nudge behavior. What begins as access to data becomes a model of a person, and then a system that acts on that model.

Privacy is no longer just about who holds your data. It is about who decides what kind of person an AI system makes of you, and what that system is permitted to do with that version. Governing this *statistical me* -- the model-generated portrait an AI assembles from traces, inferences, and interactions -- requires answers to three questions, in this order:

1. **Who has my data?** (data governance)
2. **What does the system “think” I am?** (representation governance)
3. **What is it allowed to do with me?** (action governance)

For decades, privacy practice has answered only the first. That answer remains necessary; it is no longer sufficient. The argument of this article is that the second and third questions are now privacy questions in the same sense as the first, and that AI systems have made them unavoidable.

This article builds on a recent scholarly conversation on privacy governance. Xu and Dinev (2022) argued that advances in machine learning have pushed privacy concerns beyond the flow of personal data toward what can be learned from data, shifting the center of the privacy problem from data flows to the inference of knowledge. Roemmich, Martin, and Schaub (2026) extended that move to representational inferences within models -- the embeddings, latent features, and inferred attributes that flow under the surface of system behavior. We argue that even representations are not the endpoint. As illustrated in Figure 1, privacy must now reach the *actions* AI systems take based on those representations: the routing, ranking, framing, and steering they perform once the *statistical me* has been built.

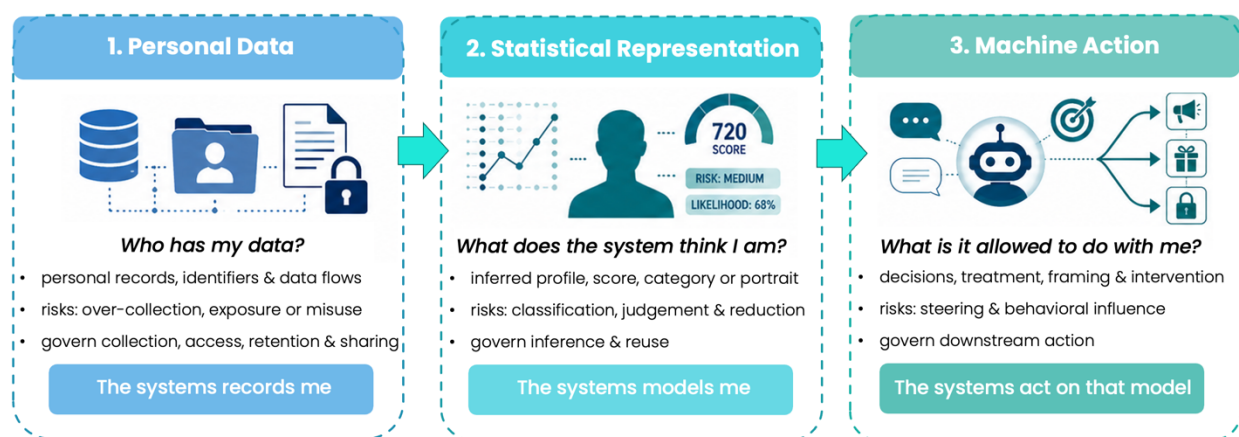


Figure 1. Three escalating privacy questions in the age of AI. *Who has my data?* captures the first-wave concern with personal data and access. *What does the system think I am?* marks the shift to machine judgment and the emergence of the statistical me. *What is it allowed to do with me?* captures the move from representation to strategic machine action. The sections that follow address each question in turn.

Who Has My Data? — and Why the Answer Is Not Enough

For decades, privacy engineering and regulation have concentrated on data and data flow (which we call the *first-wave* privacy governance): collection, notice, consent, retention, sharing, and breach. That work produced techniques -- data minimization, data anonymization, access control, encryption, auditability, impact assessments -- that organizations now depend on. It also moved the field beyond the simplistic public/private binary to a richer account of appropriate information flow within specific contexts (Malkin, 2022).

Return to the copilot example. An organization may scope its access tightly, encrypt its logs, set short retention windows, prohibit external transfer, and document every data flow. By a *first-wave* account, that looks like responsible data management. And yet, inside the system, the copilot may already have accumulated enough signals to identify who hesitates before responding, whose language suggests uncertainty, who influences whom, and who appears “high potential” or “high risk.” None of this requires unauthorized disclosure. None of it depends on transferring records to outside parties. The

system has not failed at data governance. It has simply done what AI systems do best: turned scattered traces into a coherent representation -- in this case, a model of a person.

That is the limit of *first-wave* governance. Its core logic is to discipline data flows; its core question is who gets access to records and under what conditions. Contemporary AI systems can satisfy that test in full and still construct objects -- portraits, scores, behavioral signals -- that the framework was never designed to govern. The harm is no longer only in the data handled. It is in the version of the person the system makes from that data.

What Does the System Think I Am?

The *statistical me* is the system's working theory of *you*. It is the stored profile, inferred attribute, behavioral category, or risk score that the system around the model assembles from your traces and persists across interactions. It is not the lived person, and it does not need to be accurate in any rich human sense to become consequential. It only needs to be *actionable*: good enough to route a case, raise a flag, personalize an output, prioritize a request, or quietly alter the options presented to you.

This basic phenomenon is not entirely new. Haggerty and Ericson (2000)'s data double, Cheney-Lippold (2011)'s algorithmic identity, and Citron and Pasquale (2014)'s scored society all captured the move from the embodied person to a proxy representation that can be used to classify, rank, and govern. But the *statistical me* is not simply a better proxy. The older data double was typically built for a purpose: someone chose which attributes to extract, how to score them, and where to use them. The *statistical me* is different because it is no longer confined to a particular evaluative task. In generative and agentic AI systems, it becomes the standing condition of interaction itself. The system reads the user through memory, retrieval, prior turns, and inferred signals, and then responds through that same portrait -- often opaquely, continuously, and across nearly every interaction. The *statistical me* is therefore not just a representation used by the system. It is the operational portrait through which the system knows, addresses, and acts on the person.

And in many organizations, actionability is also where the evaluation stops. Whether a representation is valid -- whether it tracks anything real about the person -- is rarely something the organization can measure in production, and so rarely something it does. The working criterion is often an operational one: is the signal discriminative enough to make different decisions about different people, with some anecdotal cases that "make sense"? Once that bar is cleared, the representation tends to stay in place, accurate or not, as nothing in the workflow is set up to discover otherwise.

To be precise: the *statistical me* is rarely an artifact of the underlying model's weights, which in most contemporary LLMs are frozen at deployment. It lives in the surrounding system: stored conversation summaries, memory entries, retrieval indices over personal documents, user-preference profile, and the data curated for the next round of training or fine-tuning. That distinction matters for governance: each of those surfaces is engineered, observable, and addressable, even when the model itself is a black

box. Governing the *statistical me* does not necessarily require interpretability of the model itself. But it does require control over the layers that construct, persist, and reuse representations of the person.

The *statistical me* has a different normative character from the source data it was built from. A calendar entry, a chat thread, a help desk interaction may each look mundane in isolation. The inference that you are disengaged, that you respond best to direct framing, or that you tend to escalate when stressed is not. These inferences travel. Once produced, they shape how later inputs are read and how later decisions are made. They can persist long after the underlying records have been deleted, and they can be reused across contexts the original collection never anticipated.

Inside the copilot, this might mean that an internal signal of “responsiveness,” learned from email cadence and meeting transcripts, becomes a feature in promotion-review tooling six months later. Nothing has been disclosed. No record has been transferred. But a representation of the person, learned in one context, has begun governing decisions in another.

This is what makes representation a privacy concern, not only a fairness one. Fairness asks whether the inference is biased. Accountability asks whether it is explainable and contestable. Privacy asks something *prior*: should this person have been rendered in this way, from these traces, for these purposes, in the first place? A model can pass the fairness test and still answer the privacy question badly. A model can be technically “private” -- every datum encrypted, every flow logged -- and still produce a representation that is profoundly invasive.

The *first-wave* privacy governance question was: *Who has my data?* The *next-wave* questions are harder: *What does the system think I am?* and *What is it allowed to do with me?* The *next-wave* concerns representation and action: how a system constructs an operational portrait from one’s traces, and how it uses that portrait to shape judgement, decisions, opportunities, and treatment. Those questions reach beyond transparency into contestability, and beyond notice into the institutional power not only to define a person operationally, but to act on that definition.

What Is It Allowed to Do with Me?

A system that can model a person can also act on that person. Earlier digital systems profiled, ranked, and targeted, but generative and agentic AI changes the texture of the problem. The *statistical me* is no longer a back-end artifact used to sort records. It becomes an active mediator between the person and the world: the layer through which information is presented, choices are framed, options are surfaced, and tasks are routed.

Return one more time to the copilot. Once the system has accumulated signals about who responds quickly, what tone secures compliance, when a worker appears overloaded, or which tasks tend to be ignored, it can do more than model those patterns. It can use them to, for example, reorder reminders, frame a recommendation in the wording most likely to persuade, escalate certain tasks and suppress

others, prefill approvals, or route interactions in ways that steer conduct without ever announcing the steering. What looks like personalization may become a form of behavioral governance.

That is a different problem from disclosure, and a different problem from accuracy or bias alone. Even a well-calibrated, non-leaking system can be invasive if it uses learned representations to shape a person's informational environment, lower the friction on some paths and raise it on others, or quietly tune what a user sees and how they are addressed. The issue is no longer only observation. It is intervention.

This argument does not depend on speculative visions of AI “scheming” (Carlsmith, 2023). Privacy risk does not require a system plotting against a user. It is enough that systems optimize *adaptively*. Once a system retains persistent signals from prior interactions and condition its outputs on them, the line between assistance and strategic influence becomes thin. A system that can predict when a user hesitates, what wording lowers resistance, or which defaults are likely to be accepted has acquired a form of asymmetrical power. *It does not need to deceive to become unsettling*. It only needs to become good at using what it has learned.

This is where current privacy governance is thinnest. Many organizations now inventory training data, run privacy impact assessments, and test models for bias. Far fewer have clear rules about when an inferred user model may drive adaptive persuasion, when personalization becomes manipulation, or when an internal assistant may act on a worker's behalf using a profile the worker cannot meaningfully inspect. Yet these are precisely the questions agentic AI forces to the surface.

Three Layers of Governance

The three questions correspond to three layers of governance. They are cumulative rather than sequential, and each addresses a distinct object of risk.

Data governance (Who has my data?) remains the foundation. Collection limits, access controls, retention, encryption, auditability, lineage, and purpose limitation do not become less important in the AI era. They become more important because everything downstream depends on disciplined upstream stewardship. But these controls answer only the first of the three questions.

Representation governance (What does the system think I am?) is the layer most current governance only partially recognize. Questions that belong to this layer include: which inferences may a system generate? Which inferred attributes may be stabilized, stored, or reused? Under what conditions may representations learned in one context be carried into another? When should an inferred representation be treated as sensitive even if the source data did not appear sensitive on its own? It also requires a different form of accountability. Documentation should cover consequential inferences and persistent summaries, not only input datasets. Organizations should provide mechanisms for contesting machine-generated portraits, not only for correcting source records. More important, there have to be

explicit limits on representational scope creep, so a system built to assist with scheduling does not quietly become a tool for evaluating reliability (or promotability) simply because the inference is technically available.

Action governance (What is it allowed to do with me?) asks how representations may be used. Does the system personalize choices in ways that steer behavior? Does it change ranking, routing, scrutiny, or friction based on inferred traits? Does it learn which framings, timings, or defaults are most likely to secure compliance from a particular user? Can it initiate tasks, recommendations, or interventions based on those representations? At this layer, governance becomes a question of power: can systems use person-models to influence, prioritize, nudge, or act on people in ways that are difficult to observe and difficult to contest.

Who should govern this? Security teams already govern access and boundaries. Privacy teams already evaluate purpose and appropriateness attached to information flows. AI governance teams increasingly assess fairness, robustness, and oversight. The most consequential harms now cut across all three. A representation learned from innocuous enterprise telemetry can become a privacy problem, a fairness problem, and a governance problem at once when it is reused to evaluate workers or steer their behavior. Treating those domains as separate is no longer sustainable.

Nor can this be governed at the organizational level alone. The three layers of governance has an institutional implication as well: the first question (who has my data?) maps relatively well onto existing data-protection structures, but the second and third questions (what does the system think I am, and what is it allowed to do with me?) put pressure on institutional arrangements built primarily around personal data and downstream system risk. To be sure, the *statistical me* will often remain in the deployer's surrounding system -- in memory layers, retrieval indices, user profiles, and other application-level components -- rather than in the frontier model itself. But where organizations rely on external model providers or managed AI infrastructure, those providers may still shape what can be remembered, inferred, audited, deleted, or operationalized. The issue is therefore not simply who "has" the *statistical me*, but how governance over representation and action is allocated across deployers, infrastructure providers, and public authorities. That allocation problem connects privacy governance directly to current debates about platform asymmetry, infrastructure dependence, and the concentration of action-shaping power in a small number of providers.

What the Law Already Sees – and What It Still Misses

Current European law already reflects part of the shift this article describes, but only part of it. The General Data Protection Regulation (GDPR)¹ remains the clearest expression of *first-wave* privacy governance: it governs personal data and processing through familiar principles of lawfulness, fairness

¹ European Parliament and Council of the European Union. (2016). Regulation (EU) 2016/679: General Data Protection Regulation. Available online at: <http://data.europa.eu/eli/reg/2016/679/oj>

and transparency, purpose limitation, data minimization, storage limitation, integrity/confidentiality, and accountability, while also reaching profiling, high-risk processing, data protection impact assessments, and certain solely automated decisions.

The European Union (EU)'s AI Act² moves further by regulating some downstream AI risks. It prohibits certain practices outright – including harmful manipulation and deception, exploitation of vulnerabilities, social scoring, and emotion recognition in workplaces and education – and imposes obligations of risk management, documentation, logging, human oversight, and robustness on high-risk systems. Even so, neither the GDPR nor the AI Act yet provides a fully developed legal framework for the *statistical me*. EU's most advanced laws still stop short of governing the core phenomena at issue here: persistent inferred representations, cross-context reuse of person-models, and adaptive action based on those models. This gap is not only doctrinal but institutional: it raises the question whether existing authorities are equipped to govern representations and actions that may be distributed across deployers, providers, and shared AI infrastructure. The harder questions still lie ahead: When should inferred representations themselves count as sensitive? What rights should people have to inspect, challenge, or confine a system's persistent summary of them? When may a model reuse a person-representation learned in one context for another? And when does adaptive personalization become impermissible steering? Future privacy laws will need to complement rights in data with rights in representations and limits on action: rights to access and contest consequential inferred profiles, rules against representational reuse without renewed justification, and constraints on agentic AI systems that use person-models to influence, prioritize, or act on behalf of individuals.

Privacy's Next Frontier

Privacy practice has outgrown the assumption that data governance is the whole problem. That assumption made sense in an earlier era, when the central privacy risks were collection without consent, excessive retention, unauthorized access, secondary use, and disclosure beyond the original context. Those remain serious problems. But AI systems now make a different risk visible and operational: the construction of statistical representations of persons that can be used to judge, sort, influence, and act on them.

The point of this article is not that personal data no longer matters in privacy discussions. It is that personal data are no longer the only, or even always the most consequential, object of privacy concern. The deeper issue is what AI systems make from data and what they are then permitted to do with what they have made. If the *first-wave* of privacy governance asked, *Who has my data?*, the *next-wave* must ask, *What does the system think I am?* and *What is it allowed to do with me?*

² European Parliament and Council of the European Union. (2024). *Regulation (EU) 2024/1689: Artificial Intelligence Act*. Available online at: <http://data.europa.eu/eli/reg/2024/1689/oj>

These three questions represent three layers of governance centered on the *statistical me*. Most organizations are currently governing only the first. That is the gap privacy researchers and practitioners now have to close.

References

- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33, 1877-1901.
- Carlsmith, J. (2023). Scheming AIs: Will AIs fake alignment during training in order to get power?. *arXiv preprint arXiv:2311.08379*.
- Cheney-Lippold, John. (2011). "A New Algorithmic Identity: Soft Biopolitics and the Modulation of Control." *Theory, Culture & Society*, 28 (6), 164–181. <https://doi.org/10.1177/0263276411424420>
- Citron, Danielle Keats, and Frank Pasquale. (2014). "The Scored Society: Due Process for Automated Predictions." *Washington Law Review*, 89 (1), 1–34.
- Haggerty, Kevin D., and Richard V. Ericson. (2000). "The Surveillant Assemblage." *The British Journal of Sociology*, 51 (4): 605–622. <https://doi.org/10.1080/00071310020015280>
- Malkin, N. (2022). Contextual Integrity, Explained: A More Usable Privacy Definition. *IEEE Security & Privacy*, 21(1), 58–65.
- Roemmich, K., Martin, K., & Schaub, F. (2026). CA–CI: Integrating Contextual Integrity and the Capabilities Approach for Dignity Considerations in AI Governance. *IEEE Security & Privacy*. <https://doi.ieeecomputersociety.org/10.1109/MSEC.2026.3654404>
- Xu, H., & Dinev, T. (2022). Reflections on the 2021 Impact Award: Why Privacy Still Matters. *MIS Quarterly*, 46(4), xx–xxxii.

Acknowledgement

This work was supported in part by the Notre Dame-IBM Tech Ethics Lab, and by the National Science Foundation under grant No. CNS2537027. Such support does not constitute endorsement by the sponsors of the views expressed in this publication.

Bio

Heng Xu is a professor at the University of Notre Dame's Mendoza College of Business, at Notre Dame, IN 46556, USA. Her research interests include data privacy, digital ethics, and fairness in machine learning, with interdisciplinary work published across information systems, computer science, law, and psychology. She is a senior member of IEEE. Contact her at heng.xu.dr@nd.edu.