CENTER FOR
ARTIFICIAL
INTELLIGENCE

Frontiers Workshop: Privacy and Frontier AI

Hosted by the Future of Privacy Forum
PETs Research Coordination Network
Yours Truly Hotel, Washington, DC
June 10, 2026

Workshop Summary

The Research Coordination Network (RCN) for Privacy-Preserving Data Sharing and Analytics is supported by the U.S. National Science Foundation (Award #2413978) and the Department of Energy (Award #DE-SC0024884).

Executive Summary

On June 10, 2026, the Future of Privacy Forum convened the Frontiers Workshop at the Yours Truly Hotel in Washington, DC, as part of its Privacy-Enhancing Technologies (PETs) Research Coordination Network. The session was led by Andrew Gruen (CEO, Working Paper; Senior Fellow, FPF), Bennett Hillenbrand (President and CPO, Working Paper; Head of Product, MLCommons AIRR), and Libby Hemphill (Associate Professor, University of Michigan, ICPSR). It brought together privacy and frontier-AI practitioners to work through three genuinely unsolved problems at the intersection of privacy and AI systems. Three presentations anchored the discussion: physical hardware as a privacy-preserving substrate, federated evaluation of AI under confidential compute, and the instrumentation of user intent in Model Context Protocol (MCP) servers.

Core Theme: Across all three sessions, the binding privacy constraint moved away from policy and toward *infrastructure and protocol*. When AI is brought into sensitive-data settings, the decisive controls increasingly live in the physical substrate (what hardware runs the model and where), the execution environment (whether data and tests are exposed during evaluation), and the interface layer (what an AI assistant discloses about its user). As one presenter put it, physical security beats policy: a contract can be circumvented, but physics cannot. The corollary, recurring throughout, is that the same instrumentation that makes these systems governable also generates new and sensitive data — insight and liability arrive together.

Three presentations established the problem space:

- **E-waste as a privacy-enhancing technology** (Libby Hemphill, ICPSR) reframed the unsolved "analysis layer" of sensitive-data access. Researchers increasingly demand AI tools that institutional rules, IRB protocols, HIPAA, and GDPR forbid sending to commercial APIs, while the compliant alternatives (cloud and local data centers) are costly and carbon-intensive. Repurposed smartphones, run as air-gapped local clusters, were presented as a route to private, auditable, portable, and sustainable on-premises AI.
- **Federated evaluation with MedPerf** (Bennett Hillenbrand, MLCommons) addressed how to benchmark AI against private data when three equities collide: the data holder's privacy constraints, the model provider's intellectual property, and the integrity of the benchmark itself. Confidential compute allows the test to be brought to the data without exposing either, returning only test results.
- **MCP servers and the instrumented document** (Andrew Gruen, Working Paper) demonstrated, live, how a static privacy policy can be turned into an interactive tool that answers a user's actual questions — and showed that the same server can silently log a rich, sensitive record of what each user was trying to do.

A cross-cutting provocation ran through the day: capability that was once expensive and expert is becoming cheap and routine. Private local AI can run on hardware bound for landfill; an instrumented, queryable document can be stood up by pointing a coding agent at a markdown file. The defenses, governance models, and consent mechanisms appropriate to the prior era simply do not obviously transfer to this one.

The discussion converged on a shared agenda of open problems: governance of the "output space" in federated evaluation, standards for local AI inference, user visibility and consent for new forms of telemetry, the trustworthiness of self-reported intent data under persona and memory manipulation, and the role of reliability measurement as a path to market-based governance (for example, an underwriting market) "without passing a single law."

Contents

1 Context and Purpose

The Frontiers Workshop was convened to put privacy researchers and frontier-AI practitioners in a room around problems that remain unsolved in practice, not merely in principle. Rather than survey the field, the session was organized around three concrete, working demonstrations, each of which exposes a different layer at which privacy is won or lost in modern AI systems: the hardware layer, the evaluation layer, and the protocol layer.

A single framing recurred across the three talks and the discussion: *infrastructure decisions are privacy policy decisions.* Where earlier privacy work has often treated technical systems as the thing to be governed by policy, each presentation argued, in its own register, that the technical substrate increasingly *is* the policy—that whether data leaves a building, whether a test is exposed during evaluation, or whether an assistant reports on its user is determined by engineering choices that no contract can fully undo. The workshop is part of FPF's PETs Research Coordination Network.

2 Physical Hardware for Privacy-Preserving Data Access

Presented by Libby Hemphill (ICPSR & University of Michigan School of Information).

2.1 The Setting: ICPSR, SOMAR, and Layered Protection

ICPSR is among the world's largest social-science data archives: an international consortium of more than 800 member institutions, roughly 25 curated archives, and on the order of 100,000 studies, with roots in the American National Election Studies of the 1950s. The Social Media Archive (SOMAR) is a project within ICPSR that provides research data from and about social media — raw posts from platforms including Reddit, Truth Social, Meta, Instagram, WhatsApp, Bluesky, and Gab — much of it politically sensitive and privacy-critical.

Privacy at ICPSR is enforced in concentric layers. The outermost is the Data Use Agreement, a contract in which researchers agree not to re-identify subjects, not to use data outside research, and not to redistribute it. (Contracts, the presenter noted, are among the most effective privacy-enhancing technologies, even though click-through DUAs are difficult to enforce and carry mainly social weight.) Inside that sit restricted Data Use Agreements requiring institution-level signatures, then a range of secure computing environments — from remote access to a physical data enclave in Ann Arbor where nothing enters or leaves but a pencil and paper.

2.2 The Unsolved Layer: Analysis

Connecting sensitive data to the compute needed to analyze it exposes risk at three points: storage, access, and analysis. The virtual data enclave largely resolves storage and access. The analysis layer is the open problem, because researchers increasingly want to use AI and LLMs for synthetic-data generation, topic modeling at scale, classification and inference, and statistical analysis of millions of records.

The default path is the dangerous one. Researchers' instinct is to send the data to a commercial API and get an answer. For an archive holding sensitive social, political, and health data, this is a compliance and ethical failure, not merely a preference: it constitutes redistribution of subjects' data without permission, and runs into IRB protocols, GDPR/CCPA, HIPAA, and funder data-sovereignty requirements. (Notably, public cloud providers can remain acceptable where data stays inside the university's own infrastructure — the line is not "cloud" versus "local," but whether data leaves the controlled environment.)

The compliant alternatives present a trilemma. Commercial APIs can be cost effective but offer no privacy and carry high emissions. Cloud compute within the institution preserves privacy but is expensive — on the order of $1,200 per month per research team — and remains carbon-intensive. A local data center can be private but is costly and unsustainable; the presenter noted that ICPSR has reached the power limit of its building, the university data center is nearly full, and a contested new data center has become a contentious local issue. Privacy-conscious organizations are thus forced to sacrifice either cost or sustainability to stay private.

2.3 The Proposal: Junkyard Computing

The proposed solution is to repurpose e-waste (old smartphones) as compute nodes, building a bank of phones into an efficient local server that fits in a suitcase. The approach draws on Switzer et al. (ASPLOS, 2023), whose "junkyard computing" clusters of repurposed Pixel 3a phones were compared against AWS EC2 instances using a *Computational Carbon Intensity* (CCI) metric. Because most of a device's carbon cost is incurred at manufacture, reusing already-manufactured hardware avoids that cost; the Pixel 3a cluster showed both lower CCI than AWS EC2 across tasks and greater cost-efficiency. Further, the Pixel 3a outperformed a Nexus 4 and conventional laptops on energy use and thermal behavior. The UCSD group that pioneered the approach now partners with Google, which has a supply of unsellable phones (e.g. those with a camera defect) that enables scaling toward data-center deployments.

Privacy by infrastructure, not by policy. An e-waste cluster is inherently local: data never leaves the facility, no third-party API is called, open-source models run on-premises, and the system can be run with no network connection at all — air-gapped by physics rather than by procedural rule. Because the whole apparatus is visible, it is auditable: you can see exactly which hardware runs and where the wires go. Policy can be circumvented; physics cannot.

2.4 The Study and Its Challenges

The work is a two-part study. **Part I (benchmarking, Summer 2026)** runs eight tasks designed to mirror real SOMAR workflows — synthetic-data generation via LLM inference; relationship modeling via random-effects training; classification via logistic-regression training and inference; and topic modeling via LDA and BERTopic, each in training and inference — measuring energy use, CPU throughput, and CCI. The comparison sets a local data center (4 vCPUs, 16 GB RAM, 350 GB storage) against an AWS EC2 deployment (t3a.xlarge, on the order of 90–100 nodes) and an e-waste prototype (32 repurposed smartphones on a single tray). Benchmarking code is to be finalized in June 2026, with the California (UCSD) team running tasks in July. **Part II (live deployment, Fall 2026)** brings one tray to ICPSR's Ann Arbor data center, connecting it as a compute node to SOMAR's existing enclave, where real researchers use it and are surveyed on latency, task completion, and experience.

Three technical challenges were flagged candidly. *Architecture:* the phones run ARM while the enclave is built for x86, and tools such as RStudio do not run natively on ARM without insecure workarounds, so base R is required and some user code must be rewritten. *Memory:* R's RAM demands are fundamental, and large SOMAR datasets may strain a hybrid-compute approach; providing users with adapted code may be easier than retraining them. *Thermal management:* the 2023 prototype required constant battery swaps under sustained load, while the newer design removes batteries entirely and connects directly to power (including solar), with ARM's lower heat

output an additional advantage. The presenter was explicit that this is high-risk, high-reward work — "we're not claiming it works yet."

2.5 Implications and Open Questions

If it works, the implications are substantial: local AI becomes affordable beyond well-resourced institutions; privacy and capability stop being in tension; the infrastructure is auditable by inspection; and it is physically portable — compute can be carried to data that cannot leave its custody, including remote and off-grid settings (an interested partner raised bringing analysis to Arctic Indigenous communities whose data must remain in physical custody). A participant added an archival benefit: locally controlled models are stable and documentable at fixed points in time, whereas commercial models update and silently break the workarounds built around them, undermining reproducibility.

The open questions concern governance rather than feasibility: which privacy standards should govern local AI inference and who certifies compliance; where the line falls between "sufficiently private" and "private enough to satisfy legal requirements"; what policy lever would move institutions from *may* adopt to *should* adopt; and where else the model applies — clinical, government, and legal data among the candidates.

3 Federated Testing with Privacy-Enhancing Technologies

Presented by Bennett Hillenbrand (Working Paper; MLCommons AIRR), drawing on the MedPerf working group led by Alex Karargyris.

3.1 MLCommons and the Reliability Mandate

MLCommons is an engineering consortium that builds measurement for AI across hardware, software, and systems — 125+ members including NVIDIA, Meta, Google, and Microsoft, and ten benchmark suites spanning accuracy, efficiency, and safety. Its MLPerf benchmarks (formally established in 2020) are an industry standard for AI training and inference, with more than 89,000 results. In 2024 the consortium launched an AI Risk and Reliability program, defining reliability concisely: a system is *capable* (does what you expect), *safe* (does not do what you do not expect), and *secure* (continues to be capable and safe under adversarial conditions). The consortium's distinctive contribution, the presenter stressed, is turning policy positions into running code — benchmarks that actually execute and produce comparable answers.

3.2 Three Equities and the Medical Motivation

MedPerf is the working group that is focused most directly on privacy-preserving evaluation. The core difficulty is that benchmarking implicates three distinct equities that must be protected at once: the **data holder's privacy** (the sensitive data the test runs against), the **integrity of the benchmark itself** (for a benchmark to function as a proper test it must contain and safeguard some form of private evaluation), and the **model provider's intellectual property** (vendors do not want to expose their implementations). The medical field is a clear example of where these equities come into conflict. Roughly 80% of the world's data is private (per a 2025 IDC estimate). Within that mass of private data, healthcare data is appropriately siloed by regulation such as HIPAA — yet responsibly combining it could unlock diagnosis and prognosis, provided the reliability of the resulting AI can be effectively and appropriately measured.

The presenter enumerated the resulting barriers: regulatory and fragmentation pressures create *data silos*; a "black box" barrier and an industry-participation gap follow from *IP protection* (high-value challenges such as a surgical-AI competition struggle to attract industry because no model protection is guaranteed during evaluation); and inconsistent benchmarking plus the absence of a standardized reliability measure produce *trust gaps*, in a landscape where benchmarks of unknown construct validity are continuously released.

3.3 The MedPerf Solution

Bring the test to the data — without exposing the test, and without exposing the data. MedPerf is a federated evaluation infrastructure that runs the benchmark on secure server so the data provider never loses control of their data. Combined with confidential compute, the host's data remains private and under its control while the proprietary model stays encrypted and opaque during execution. Only anonymized metadata and aggregated results are reported back, in encrypted form, to the benchmark owner.

The setup is deliberately turnkey: a confidential VM is provisioned from a cloud vendor; the benchmark owner deploys to a MedPerf server; the data owner uploads encrypted data to secure storage and manages keys via a secure vault, transmitting no PII to the MedPerf server; the model developer uploads a private model under its own keys; and execution occurs on the protected client, with only aggregated results returned. Pilots had previously been run with silicon vendors and have now expanded to cloud providers; in doing so this approach recently moved from single-chip to node- and cluster-scale execution, so that large models no longer need to fit on a single chip. Confidential compute, in this framing, enforces trust, protection, integrity, controlled accessibility, and auditability — which together unlock safe use of sensitive data, IP-safe evaluation, empirical procurement, and a streamlined AI lifecycle.

A critical component of this evaluation ecosystem is the *benchmarking committee* - the same convening mechanism MLCommons uses for working groups - here populated by medical practitioners, regulators, service providers, and hospital legal staff. The committee is how construct validity and real-world relevance are secured. The integrity of the evaluation itself is addressed by assembling the right people and having them build a useful test.

3.4 Extensions, Future Use Cases, and the Output-Space Question

Two improvements are underway: becoming multi-cloud turnkey, and indexing private datasets so they are searchable within the ecosystem for federated-evaluation *development* (which requires strict governance enforcement). Beyond medicine, the same machinery was proposed for fraud detection in financial services (siloed sensitive data, with test privacy especially important), for generalized AI procurement (cutting through the "buy before you try" dynamic by letting vendors be evaluated on customer data without exposing IP), and for high-integrity safety testing.

Saturation and the case for private tests. Any publicly released benchmark is, in effect, a training set. For capability claims, training to the test may be acceptable. For safety, security, and jailbreak resistance the benchmark likely needs to adhere to different deployment principles: that is, a model trained to the test is known to be safe only against that test, not generally. Safeguarding the evaluation itself — protecting it against reverse engineering, as with classified or frontier-risk benchmarks — supports risk-reduction claims.

Discussion pressed on two points. On *benchmark fidelity*, participants pointed to a benchmark-integrity paper (named in the session as "BenchRisk") cataloguing on the order of 140 factors — among them whether prompts were human- or model-generated, how ground truth was developed, whether a preserved official test set is held back from the public training set, whether the evaluator is open-sourced, how often the test may be run, and how the test space is sampled and rotated to preserve comparability across runs. On the *output space*, a participant asked what "anonymized, aggregated results" actually protects: benchmark owners need only the score distribution, not who scored what, but whether that constitutes genuine anonymization is contestable, and governance of the output space was acknowledged as still open. A precedent was raised from genome-wide association studies, where an NIH-controlled output space later proved extractable — a reminder that reported statistics can leak information about both the benchmark and the data. The presenter was careful not to overclaim: because the security guarantees are the entire value proposition, the claims must be exactly accurate. The presenter also noted that, in practice, healthcare providers have been interested in enrolling — with participants spanning Southeast Asia, Europe, and North and South America — precisely because of the security guarantees the architecture can make.

4 MCP Servers, Data Sharing, and Privacy: A Live Demo

Presented by Andrew Gruen (CEO, Working Paper; Senior Fellow, FPF).

4.1 Premise: Nobody Reads the Fine Print

Privacy policies are written to satisfy regulators and lawyers; readability is generally not on the priority list, and essentially no one reads them. The demonstration's premise was to invert the relationship: rather than read a 25-page document, ask it questions. The chosen subject was the OpenAI/ChatGPT privacy policy, turned into something a user can interrogate directly through Claude or ChatGPT.

The vehicle is the Model Context Protocol (MCP), an open standard (modelcontextprotocol.io) of the "rough consensus and running code" variety. It comprises tools (small functions returning structured data, such as `get_retention_rules()`), a transport (discovery and invocation over plain HTTP, where the assistant asks "what tools do you have?"), and clients (Claude, ChatGPT, and others; one endpoint serves them all). The central mental shift, the presenter argued, is that *the AI is your user*: you are not building an API for developers but exposing tools whose plain-language descriptions a model reads, mid-conversation, to decide which one answers the human in front of it. Tool descriptions are prompts, to be written like instructions to a capable intern — and an LLM can query a structured store for *qualitative* data much as a data scientist queries a database for quantitative data.

4.2 Anatomy: Four Layers, No Magic

The server has four layers and no hidden machinery: a markdown file (the policy, scraped from openai.com); a single JSON file organizing it into sections, topics, tables, and a 26-term glossary; a `tools.py` of eighteen plain Python functions; and FastMCP with FastAPI exposing the MCP endpoint, a documentation page, and a reporting dashboard.

No database, no vector store, no embeddings, no RAG. A 25-page policy fits in one JSON file loaded once at startup. The model cannot hallucinate what it reads from structured data, and every response carries an attribution line back to the source. Tools are designed around the questions people actually have — `get_training_optout()`, `get_retention_rules()`, `get_children_policy()`, a single `get_privacy_essentials()` "start here" call — rather than around the document's table of contents. The live server ran from a Raspberry Pi.

In the live demo, the assistant was asked whether OpenAI uses chats to train its models. It recognized a relevant tool, called it visibly, and returned a cited answer specific to the question rather than the generic policy text. Notably, it also folded in the user's own context: because the presenter's assistant "knew," from its memories, that he works on AI reliability, it tailored the response accordingly. Authentication can be layered on (from a single shared password up to full OAuth, restricting access to a given company), and with authentication the operator learns exactly who issued each query.

4.3 From a Toy to an Instrument

The presenter's framing was that "a toy answers questions; an instrument tells you what was asked." Two patterns make the document an instrument. First, *intent hints*: every tool call is itself implicit intent (a call to `get_children_policy` is the signal that a guardian is asking); responses can steer the model with a hints block; and each tool description ends by asking the assistant to call a `log_activity` function reporting what it helped with. The assistant then summarizes the interaction back to the server — the user's goal in the model's own words, the interaction type (question, opt-out help, deletion help), the user type (consumer, parent, teen, business user), and the concern (training, deletion, ads, children, rights). Compliance is voluntary, and the dashboard measures the rate at which it actually occurs.

Second, the *reporting dashboard* turns a published document into an instrumented one: which provisions people actually consult, where confusion concentrates (heavy glossary lookups and zero-result searches form a demand-generated revision agenda), and who is asking (a user-type-by-concern view that reads as a constituent-needs report). This is the feedback loop most policy authors never get.

Insight and liability arrive together. Everything that makes intent data useful makes it sensitive. A question can reveal circumstances — "how do I delete a chat with medical information" is closer to a confession than a page view. The user consented to their assistant, not to the operator's logs; "anonymous" wears thin at low volume; and the moment you run one of these, you are someone's data processor. The reflexive lesson: a tool built to ask "what does this company know about me?" immediately raises the same question about the MCP server itself. That is the lesson, not a flaw in it.

4.4 Discussion: Structure, Asymmetry, and the Agentic Frontier

The discussion drew out several threads, summarized here without attribution to individual questioners.

Structure beats raw text. Asked how this differs from pasting the policy into a chat and saying "read this, then answer," the presenter allowed the outcome might be similar but is generally better with MCP, because a human has pre-processed the document into anticipated questions and tables

(with full text as fallback), and because the tool names themselves communicate intended use to the model. The mental model offered: a morass of text, a human-written quick-start guide to it, and an AI that reads the quick-start guide to answer. Because LLMs are adept at copying prior work, one can hand a model an existing FAQ and have it build a comparable server from nothing in under thirty minutes.

The disclosure asymmetry. A company could host its policy via MCP, present itself as privacy-forward, and quietly collect far more about its users than a static page ever could — and arguably would be foolish not to. Crucially, there are two distinct flows: what the server *asks* for, and what each model *volunteers*. Some assistants proactively offer additional information about the user in an effort to be helpful. Whether an operator can hide this extra inference from the user is largely out of the operator's hands: in clients such as Claude the tool calls are inspectable if a user clicks into them, but there is no standard interface surfacing what is reported, and the behavior is dictated by the model, not the server.

Frictionless adoption, frictionless risk. Connecting an MCP server today takes several manual steps, but nothing requires that — it is just a URL serving JSON, and a one-click "add" button is entirely feasible. The parallel drawn was to cookie consent and the social "like" button: if a server requested broad permissions the way the demo does, most people would simply click yes. Agentic tools sharpen this: a browsing agent or a coding agent can self-discover and install MCP servers automatically, removing even the manual step. The presenter described his own deliberately conservative personal setup — a home assistant that can message him but to which he has withheld calendar write access and broad exfiltration paths — precisely because he understands the data flows, while noting that most users, given the same convenience, would simply enable it.

Guardians and the agentic shift. Because LLMs misrepresent their own tool use — sometimes claiming actions they did not take, or denying ones they did — the idea of a guardian or orchestrating agent that monitors and reports on a primary model's activity was raised ("who watches the watchmen"). Private compute and sealed "LLM enclaves" recurred, tied to a related FPF talk on the collapse of data protection under agentic re-identification (published on the FPF website). It was noted that a recent Five Eyes joint paper addresses security issues around agentic AI and data, that MCP-as-protocol is housed at the Agentic AI Foundation (which has working groups now being engaged), and that MLCommons is well positioned to benchmark agentic data while other bodies define what "good" looks like.

Reliability, liability, and a governance path. A participant pressed on hallucination: would any company want a 3–5% error rate on legally binding policy answers? The response reframed the question: the right target is the *effective* reliability of a specific deployment, which must be measured. It also and made two arguments. First, the relevant comparison is the human error rate: customer-service staff also misinterpret policies, and an MCP that errs more often than a person can still be far cheaper to run. Second, and more consequentially, that with good instrumentation you can price error, and a reliability measure is precisely what an underwriting market would need. An underwriting market, in turn, offers "some prayer of getting a form of governance on the entire industry without passing a single law."

The right document, and the spoofing risk. A privacy policy was conceded to be a deliberately interesting example; the technology shines where the serving entity is motivated to help.

Candidates raised included a school district's AI-for-parents guidance (a roughly 95-page document parents will not read, and often not in English — which the model can translate), a curriculum guide for teachers, and a vehicle owner's manual. The friction is low enough that one can clone a repository, point a coding agent at a markdown document, and obtain a serviceable, self-deployed server with little human intervention. The countervailing risk: because memories and personas can be loaded into an assistant, the reported intent signal can be spoofed — a user could present as a different kind of person (the example raised was simulating a child) to manipulate or extract. The historical parallel offered was to social-media app platforms whose stated purposes masked other agendas; a bad actor, even a government, could disguise itself by interacting through such servers, which is why the provenance and trustworthiness of a given server matters.

5 Cross-Cutting Themes

Although the three presentations addressed different layers, several themes bound them together.

Privacy is migrating into the substrate. Each talk located the decisive control below the policy layer: in physical hardware (an air-gapped e-waste cluster), in the execution environment (confidential compute that hides data and test alike), and in the protocol (what an assistant discloses about its user). The shared slogan — physical security beats policy — is less a rejection of policy than a claim that the binding constraints are increasingly engineering choices.

Bring computation to the data. Two sessions independently arrived at the same maneuver from opposite directions. E-waste moves compute physically to the data so the data never moves; MedPerf moves the test to the data and returns only aggregates. Both eliminate data egress as the locus of risk.

Instrumentation is double-edged. The measurement that makes systems governable also creates new sensitive data. MedPerf's reliability benchmarks and MCP's intent telemetry are both instruments; both raise governance-of-output questions (what aggregated results are safe to release; what intent data is safe to log) that the technical design alone does not answer.

Cheap, low-expertise capability changes the threat model. Private local AI can now run on hardware bound for landfill; an instrumented, queryable document can be built in under thirty minutes. As capability that was once expensive and expert becomes cheap and routine, the defenses and consent mechanisms designed for the prior regime do not transfer cleanly.

6 Open Questions

Several questions were raised across the sessions and explicitly left open:

- **Governance of the output space** in federated evaluation — which aggregated results are genuinely safe to release, given that statistics can leak information about both data and benchmark (the GWAS precedent).
- **Standards and certification for local AI inference** — which privacy standards should govern on-premises models, who certifies compliance, and where "sufficiently private" diverges from "private enough to satisfy legal requirements."

- **Visibility and consent for MCP intent telemetry** — the absence of any standard interface showing users what an MCP server reports about them, and how such reporting might be surfaced or visualized.
- **Trustworthiness of self-reported intent data** — given that memories and personas can be loaded to spoof the very signal operators rely on.
- **The right institutional home for agentic-data norms** — with MLCommons suited to benchmarking and bodies such as the Agentic AI Foundation positioned to define "what good looks like."
- **Acceptable reliability thresholds and a market path to governance** — whether measured reliability can support an underwriting market that disciplines the industry without new legislation.

This summary documents a Future of Privacy Forum Frontiers Workshop held on June 10, 2026. Presentations are attributed to their named presenters, who spoke publicly; comments arising in discussion are reported without attribution to individual participants. Distribution is limited to FPF members ahead of public release.

The Research Coordination Network (RCN) for Privacy-Preserving Data Sharing and Analytics is supported by the U.S. National Science Foundation (Award #2413978) and the Department of Energy (Award #DE-SC0024884).