IIW SESSION 3 OF 4

The AI Accountability Gap

What GAAFA does that SEDI and PDTA cannot — and what the technical community needs to build to make it work

Michael G. Leahy | Internet Identity Workshop | April 2026

Format: 60 minutes | ~35 min presentation / ~20 min discussion / 5 min close

Format: 60 minutes | ~25 min presentation / ~30 min discussion / 5 min close

---

Session Purpose | Present the specific problem that AI creates for any privacy-preserving identity and data framework, introduce GAAFA's three mechanisms as the legislative response, and extract the technical community's expertise on whether those mechanisms are implementable. This is the most technically demanding session and the one where IIW participants have the most to contribute. The presentation is deliberately shorter to leave maximum room for substantive critique.

A. Opening and Continuity *[4 min]*

- One-paragraph re-orientation for new participants — SEDI governs the identity layer, PDTA governs the legal relationship to data, this session addresses what happens when AI makes the decisions from that data
- Room identification — specifically ask if anyone works directly on AI systems for government clients or on AI governance frameworks

  Facilitator note: *This room will likely skew more technical than Sessions 1 and 2. Adjust the depth of the architecture discussion in Section B accordingly.*

B. The Accountability Gap — Stated Precisely *[10 min]*

- The gap defined: the space between government's obligations regarding citizen data and its obligations when AI systems are making the operative decisions from that data
- Concrete contexts where the gap is active today
  - *Benefits eligibility determinations — Social Security, Medicaid, unemployment*
  - *Criminal risk assessment and pretrial detention recommendations*
  - *Tax audit selection and enforcement prioritization*
  - *Immigration adjudications and visa processing*
  - *Child welfare intervention decisions*
  - *Licensing and professional certification determinations*
- Why SEDI and PDTA together do not close the gap
  - *SEDI prohibits surveillance at the identity transaction layer — it does not constrain the decision made from data government already holds*
  - *PDTA imposes fiduciary duties on human secondary fiduciaries — it does not directly govern automated systems that are not human actors*
  - *An agency fully compliant with both statutes can deploy a black-box AI that makes the actual rights-affecting determination*

? Has anyone in this room built or evaluated an AI system deployed in one of these government decision contexts? What visibility did affected citizens actually have into how the decision was made?

C. The EU AI Act as a Cautionary Precedent *[8 min]*

- The EU AI Act's ambition — the world's most comprehensive attempt at AI governance, risk-based tiering, high-risk category requirements
- The two structural vulnerabilities that have delayed and weakened it
  - *Reliance on harmonized technical standards that were not developed in time and cannot be compelled by legislative deadline*
  - *Reliance on enforcement authorities that were not constituted by statutory deadlines — the governance structure preceded the governed technology*
- The lesson for American legislative design: durable AI governance must be grounded in constitutional principles and existing institutional frameworks, not in technical specifications or new agencies that depend on future legislative action
- What this means for GAAFA's design choices — and what it means for what this community needs to build

  Facilitator note: *This section will generate strong reaction from the IIW community, many of whom have followed the EU AI Act closely. Let brief reactions surface and park them for Section E.*

D. GAAFA's Three Mechanisms *[12 min]*

- Mechanism 1: Algorithmic Deployment Assessments
  - *Required before any rights-affecting AI system is deployed — not after, not as an annual audit*
  - *Structured evaluation of decisional logic, disparate impact potential, interpretability characteristics, and human oversight capacity*
  - *The open question: what technical standards should the assessment require, and who is competent to conduct it?*
- Mechanism 2: The Right to Algorithmic Explanation
  - *When AI makes or substantially influences a decision affecting a citizen's rights, that citizen is entitled to a meaningful account of how the decision was reached*
  - *In terms accessible to a non-expert — not a technical audit report*
  - *Human review by an official with actual oversight capacity — not nominal supervisory authority over a system they do not understand*
  - *The open question: what does 'actual oversight capacity' require technically, and can it be defined in statute without becoming a technical specification subject to EU AI Act-style implementation failure?*
- Mechanism 3: AI Agents as Statutory Secondary Fiduciaries
  - *Autonomous AI agents deployed by government to query, process, or act on citizen data are classified as statutory secondary fiduciaries*
  - *Subject to the same duties of loyalty, care, and transparency that PDTA imposes on human secondary fiduciaries*
  - *Black-box models in rights-affecting contexts are treated not as a technical limitation but as a breach of pre-existing fiduciary obligations*

- *The open question: does fiduciary law generate coherent obligations when applied to automated systems, or does it import duties designed for human actors that cannot be discharged by AI?*

Facilitator note: *Section D is the most technically contentious part of the entire four-session arc. State the open questions explicitly and with genuine uncertainty — this is where the room's expertise is most valuable and where you most need their critique.*

E. Discussion *[21 min]*

? The deployment assessment requirement assumes that someone can evaluate AI decisional logic before deployment. In the government AI systems you have built or evaluated, is that assumption realistic, and if not, what would make it so?

? The right to algorithmic explanation requires a meaningful account in terms accessible to a non-expert. Is that technically achievable for the categories of AI systems being deployed in high-stakes government decisions today, or does it functionally prohibit the most capable models?

? Classifying AI agents as fiduciaries imposes duties of loyalty and care. What would it mean in practice for an AI system to discharge a duty of loyalty to the citizen whose data it processes — and is that a coherent obligation or a category error?

? The EU AI Act failed in part because it depended on technical standards that could not be developed on a legislative timeline. What standards does this community have — or is building — that could make GAAFA's mechanisms implementable without that dependency?

? If interpretability is required for high-stakes government AI and current high-capability models are not fully interpretable, does the statute effectively mandate less capable models in those contexts? Is that the right tradeoff?

Facilitator note: *The last two questions are the most important for post-IIW engagement. The IIW community is actively building the technical infrastructure that makes this legislation either possible or impossible. Identify who in the room is working on interpretability, verifiable credentials for AI outputs, or AI audit standards — those are your follow-up conversations.*

F. Close and Call to Action *[5 min]*

Bridge to Session 4: We have now covered all three layers of the framework separately. Session 4 asks whether they hold together as a unified constitutional architecture — and whether that architecture is both legally robust enough to withstand doctrinal challenge and politically replicable beyond Utah.

Call to Action:

→ If you work on AI interpretability, algorithmic auditing, or verifiable credentials for AI outputs and you believe GAAFA's mechanisms are technically flawed, I am specifically requesting your written critique — contact me after this session

→ Identify the one mechanism in GAAFA that you believe is most technically achievable in the near term and the one that is least achievable — bring both to Session 4

→ If you advise state governments on AI procurement or deployment, the deployment assessment requirement is where this legislation would first touch your work — I would welcome a conversation about whether the statutory language reflects procurement reality