BCLT AI Governance at the Crossroads, Paul Ohm

1. Prep doc

AI Governance Symposium Navigating AI Frontiers: Strategies from the States Feb. 28, 2025 | Panel 2: 11:05-12:25 Location: Residence Inn, Berkeley (follow signage at entrance) Conference webpage w/ agenda

California Governor Gavin Newsom was among the first nationally to issue an executive order calling for the development of sound responsible AI governance strategies in state government. This panel features representatives from state government agencies and leading academics who discuss the ramifications of executive orders and other state activities, highlighting the proactive steps taken by states to incorporate and support AI technologies while safeguarding public interests. One focus will be on the collaborative efforts to establish robust responsible AI procurement processes, including the development and implementation of AI risk assessments and strategies to assess the credibility of third-party AI risk auditors, certifiers, and licensors. Another will be on state governance of AI outside governmental use. Panelists will address the challenges and opportunities inherent in these efforts, offering valuable insights into the practicalities of implementing and maintaining rigorous AI oversight mechanisms in state government.

Invited participants [CONFIRMED] Brandie Nonnecke (moderator), Director, CITRIS Policy Lab; Assoc. Research Professor, Goldman School of Public Policy; Co-Director, Berkeley Center for Law & Technology

[CONFIRMED] Jennifer Urban, Chair, California Privacy Protection Agency Board, Clinical Professor of Law; Director, Samuelson Law, Technology & Public Policy Clinic; Co-Director, Berkeley Center for Law & Technology

[CONFIRMED] Jonathan Porat, Chief Technology Officer, California Department of Technology

[CONFIRMED] Paul Ohm, Professor of Law, Georgetown Law

11:05 - 11:10 | Introduction of panel and speakers (Brandie Nonnecke) 11:10 - 11:15 | Jennifer Urban 11:15 - 11:20 | Jonathan Porat 11:20 - 11:25 | Paul Ohm 11:25 - 12:15 | Panel Discussion (see draft questions below) 12:15 - 12:25 | Audience Q&A

Draft Panel Questions State AI Governance Leadership

California is among the first to issue executive orders on responsible AI governance. What motivated these early actions, and how have they influenced state-level AI policies nationwide?

Procurement and Risk Assessment

Both states are working to establish responsible AI procurement processes, including AI risk assessments and oversight mechanisms for third-party auditors and certifiers. What best practices have emerged from these efforts, and what challenges still need to be addressed to ensure procurement processes prioritize public interest and accountability?

Beyond Government Use

While much focus has been on how AI is used within government, states also play a role in shaping the broader AI ecosystem. How is California approaching governance of AI outside of direct governmental use, such as through consumer protection, labor policies, or public-private partnerships?

Collaboration and Scalability

Developing responsible AI governance frameworks requires collaboration across agencies, academia, and industry. What lessons have been learned from interagency and cross-sector partnerships, and how can these efforts be scaled or adapted for other states?

Future Challenges and Opportunities

Looking ahead, what are the biggest risks and opportunities for state governments in AI governance? Are there particular areas where state leadership is needed to fill gaps in federal AI regulation, and what role do states play in setting national AI governance norms?

2. New Draft

- Trying to capture the major themes/moves
- and trying to separate my measly 5 minutes from Q&A

2.1. Opening

- I have some Colorado-specific lessons learned from my time working on the CPA rules, but I want to delay those until Q&A so I can begin to describe a new project of mine

2.1.1. The "patchwork" is a revision control system

- Making each bill better than the last one
- Laboratories in action
  - Keep what's good (allowing interop) and throw out or improve on what's bad
  - CA -> CO: Opt-out to opt-in for sensitive info
  - CO -> MD: Robust data minimization
  - Some day soon? Private right of action
- Patchworks can be beautiful and harmonious! And they can keep you warm during the winter.
  - Law firms and compliance industry will rally!
  - See data breach notification
- Interop
  - In CPA: rules about data impact assessments:
    - Rule 8.02(B): If a Controller conducts a data protection assessment for the purpose of complying with another jurisdiction’s law or regulation, the assessment shall satisfy the requirements established in this section if such data protection assessment is reasonably similar in scope and effect to the data protection assessment that would otherwise be conducted pursuant to this section.
      - 1. If a data protection assessment conducted for the purpose of complying with another

jurisdiction’s law or regulation is not similar in scope and effect to a data protection

assessment created pursuant to this section, a Controller may submit that assessment with a supplement that contains any additional information required by this jurisdiction.

- Not quite as good, also In CPA: rules about notice incorporating CA privacy policies
    - Rule 6.02(B): A Controller is not required to provide a separate Colorado-specific privacy notice or section of a privacy notice as long as the Controller's privacy notice meets all requirements of this section and makes clear that Colorado Consumers are entitled to the rights provided by C.R.S. § 6-1-1306
- Echoes of it in AI Act's provision: "6-1-1703(e) IF A DEPLOYER, OR A THIRD PARTY CONTRACTED BY THE DEPLOYER, COMPLETES AN IMPACT ASSESSMENT FOR THE PURPOSE OF COMPLYING WITH ANOTHER APPLICABLE LAW OR REGULATION, THE IMPACT ASSESSMENT SATISFIES THE REQUIREMENTS ESTABLISHED IN THIS SUBSECTION (3) IF THE IMPACT ASSESSMENT IS REASONABLY SIMILAR IN SCOPE AND EFFECT TO THE IMPACT ASSESSMENT THAT WOULD OTHERWISE BE COMPLETED PURSUANT TO THIS SUBSECTION (3)."

2.1.2. Rampant hypocrisy at work

- On the one hand, tech bros talking a Y Combinator or pitching Andreesen or Pounding their chest on X claim that AI is on the brink of having the power to do EVERYTHING!
- On the other hand, when states pass sensible, modest, and frankly, incrementalist new proposals putting lightweight obligations on companies, suddenly, they fall on the ground crying.
    - On no, don't require us to do any risk assessment!
    - On no, don't require us to add any guardrails.
- Many of these laws obligate companies to do what they are already doing, inspired by ethics and community norms and a sense of obligation to their fellow man.
    - So the prohibition isn't really about the *burden* of the law.
    - It's a deeper political objection to the source of the obligation.
- Also, you can use AI to do a lot of this compliance work!
- I get it; that's a valid debate we ought to be having: if a law requires you to do little

more than you would already be doing; if it costs you pennies on the dollar to implement; should you still disallow for deeper philosophical or political reasons?

- Perhaps, but let's not pretend it's about costs and benefits.

2.1.3. The possible impossible

- I am engaged in a multi-pronged research agenda noting something that most smart model builders understand but that most smart legal scholars and policymakers have not yet internalized.
  - Foundation models, and the applied models they spin out, are among the most plastic and malleable industrial creations since the dawn of the industrial age.
- We're just starting on the work, so let me focus on one claim: changing the fundamental nature of a foundation model is astonishingly cheap and simple.

2.1.4. Prescriptions

- One might be ironic: if you're worried most about regulatory burden, it's better to just ban behaviors and put behavioral obligations on models, rather than impose costly risk assessments and audits.
  - But those have to be backed by the threat of investigation and liability of course!
  - For example, with AI companions:
    - rather than say "a company marketing an AI companion must assess the risk that their model will encourage or instruct minors to harm themselves"
    - a law should instead say, "a company marking an AI companion must ensure that their model will not encourage or instruct minors to harm themselves"
  - Phrasing it as a prohibition, as a design instruction, might be counter-intuitively far less burdensome than requiring an impact assessment.
- But wait, you say, won't it be costly to prevent an AI companion to avoid instructing minors to harm themselves?
  - No, not at all!
  - You can fine-tune it
    - Using labeled examples

        - Using reinforcement learning with human feedback
        - Maybe even using reinforcement learning with non-human feedback
    - You won't really have to write any code
    - You just need to curate the dataset or reinforcement learning architecture, but you're already doing that to serve another 1000 goals
    - I want to leave you with a concrete estimate, not based on any scientific measurement, but based on a well-honed intuition: it'll take three engineers a week and a case of Celsius energy drink
        - I might be wrong about the preferred stimulant, but that's in the ballpark
- I'm sure there will be many libertarians in the audience clutching their pearls at this estimate.
    - Oh dear god, did he say regulation won't be burdensome? Oh my!
        - The Netchoices and the fake think tankers with ridiculous fake testosterone spiked pseudonyms like "Dean Ball".
    - I'm not really talking to you. I won't convince you, and you'll get paid by the word to object despite what evidence and common sense tell you.
- I'm talking to the regulators and the civil society advocates, and ultimately, to the courts
    - In the entire history of industrial regulation, there has never been a core set of technologies as inexpensive to fundamentally reshape than models based on foundation models.
    - We should be writing simple, straightforward laws of the form: you model must not do X, and if it does X, you are responsible.

2.2. Q&A themes

2.2.1. Checks and balances and the last bulwark we have got!

- State govt legislation is the only bulwark you have. Even if you entered the year sympathetic to concerns about the overregulation of the AI industry, I hope you've been given a little hesitation in the past month from the dangers of unchecked power in the hands of too few.
    - Forget laboratories of democracy, this is one of the last checks and balances we

have!

2.2.2. Lessons from Colorado

1. The importance of leadership
2. The importance of competition
3. The importance of smaller states getting on the field
    - True, the costs of participation are really, really difficult for civil society
    - But I detect that they are hard for tech companies too
        - We held several public meetings
        - Astonished by how few companies and trade associations there were
            - They tended to send their B teams
            - They tended not to use their best arguments
            - They clearly didn't know the proposals very well

2.2.3. The FPF episode

2.2.4. On private rights of action

- Note California's innovation: State AG, County atty, City atty (SB 942)

3. Key terminology from laws (cheatsheet)

3.1. CO

- "Makes, or is a **substantial factor in making,** a consequential decision"
- "reasonable care to protect consumers from any known or reasonable foreseeable risks of algo discrim"
- algo discrim classes: age, color, disability, ethnicity, genetic info, limited proficiency in English, national origin, race, religion, reproductive health, sex, veteran status, or other classification protected under law

3.2. IAPP's taxonomy of types of laws proposed so far

3.2.1. AI governance programs

3.2.2. Assessments (risk, impact, rights)

3.2.3. User training (not common)

3.2.4. Responsible individual

3.2.5. General notice

3.2.6. Disclosures/labeling

3.2.7. Explanations and incident reporting

3.2.8. Documentation from Dev to Deploy

3.2.9. Registration

3.2.10. Third-party rule (not common)

3.2.11. User opt-out/appeal

3.2.12. Nondiscrimination

3.3. FPF's list of recurring issues/differences

3.3.1. The Consequential Decisions (ADM) approach ("most prevalant")

1. Which contexts?
    - ed, jobs, housing, financial, govt services, health, insurance, legal
    - CA adds: utilities, CJ, adoption, repro svcs, voting
2. Which effects?
    - "legal or similarly significant effects"
    - CO adds: "cost or terms of"
    - CA adds: "impact of, availability of"
3. Role of AI system ("most debated and difficult")
    - Spectrum from most to least coverage:
        - "facilitating decision making" (CPPA)
        - "substantial factor" (CT imported into CO)
        - "controlling factor"
            - VA HB 2094: "specifically intended to autonomously make, or be a substantial factor in making"
4. Dev vs Deploy
5. Common Exceptions
    - Tech (calculators)
    - Other laws
    - Small businesses
    - Public interest–protect a human life
6. Algo discrim

    - Two main approaches
        - CA: prohibition: If reas risk of AD, cannot use tool
        - CO: rebuttable presumption to duty of care if you comply with statute

3.4. For ADM-style laws, the different options

- Participation in the decision
    - CO: A substantial factor in making
    - VA: Specifically intended to autonomously make

Author: Paul Ohm

Created: 2025-02-28 Fri 10:26

Validate