

TO: Governor Gavin Newsom

FROM: Policy Advisor, Office of Critical and Emerging Technology

DATE: April 14, 2025

RE: Advancing Risk-Based AI Governance in California Post-SB 1047

I. Overview

Governor Newsom, you vetoed SB 1047 because its focus on large-scale AI systems was an imprecise proxy for risk. Alternatively, you proposed that a more “evidence-based and adaptable” framework would be necessary, avoiding one-size-fits-all regulation that unduly limit startups.

This memo proposes a more specific and not as general framework as SB 1047. We believe strengthening oversight of high-stakes AI development (such as in housing, law enforcement, and employment) leveraging impact assessments, audits, and a regulatory sandbox will avoid arbitrary dollar triggers, align with industry best practices, and make California a contributing leader in developing ethical and effective AI governance.

II. Impact Assessment

Any AI system used by California state agencies—or third-party vendors on their behalf—in domains such as housing, policing, or employment should undergo a mandatory AI Impact Assessment (AIA) prior to deployment. These assessments will evaluate the model’s design, purpose, training data, risk of bias, and plan for mitigation or human oversight.

Rationale:

- A lawsuit in Massachusetts found that a tenant screening algorithm (SafeRent Solutions) discriminated against applicants using Section 8 vouchers, often Black and Hispanic renters, by unfairly penalizing them based on credit scores while ignoring voucher guarantees. The result was a \$2.3 million settlement and new oversight.
- Predictive policing algorithms, such as those trialed in Los Angeles and Chicago, were shown to reinforce historical bias, leading to over-policing in communities of color. Independent audits recommended their removal.
- Microsoft and Google already mandate internal AI impact reviews, especially for “high-risk use cases,” and share frameworks publicly.

These assessments build transparency, prevent harm, and ensure public trust—without interfering with low-risk AI systems or innovation pipelines.

III. Audit

AI models used in important public areas—no matter how much they cost to develop—should go through regular audits for fairness and performance, plus stress testing before launch to catch edge cases and potential failures.

Rationale:

- Many systems posing risk to civil rights are not expensive frontier models but domain-specific, fine-tuned tools. For example, resume-screening algorithms have historically favored men due to biased training data.
- NYC’s Local Law 144 requires annual bias audits of employment-related AI tools, and California could build on this precedent.

- OpenAI, Meta, and others have publicly committed to “red teaming” and safety testing prior to deployment. These practices can be localized via California-certified third-party audits.

Audits will make sure there's accountability, while stress tests will help catch potential problems before they cause harm in important situations.

IV. Regulatory Sandbox

Set up a California AI Sandbox—a safe, guided space where companies can try out new AI tools with temporary rule flexibility, especially for startups and projects that serve the public good.

Rationale:

- This mirrors successful fintech sandboxes (e.g., in Arizona and Utah), and provides a scalable model for testing without immediate compliance burdens.
- The EU AI Act, while comprehensive, has been criticized for limiting innovation. A California sandbox allows firms to “test, learn, adapt”—ensuring safety without stagnation.
- Public agencies can co-participate as partners, helping developers align early with social needs and standards.

A sandbox will connect innovation and regulation, giving startups clarity and confidence to enter public-sector markets.

V. Conclusion

This framework responds directly to concerns raised in your SB 1047 veto:

- No arbitrary dollar thresholds. Regulation is based on real-world impact and deployment context, not model size or training cost.
- Promotes innovation. By offering sandboxing and clear but flexible accountability standards, California stays competitive as a tech leader.
- Grounded in precedent. Mirrors best practices already adopted by leading tech companies and public agencies across the U.S.
- Scalable and enforceable. Builds on tools already piloted in other domains (e.g., employment law, digital services), ensuring administrative feasibility.

Revising the framework will allow California to lead with smarter, risk-based, and innovation friendly policy that responds to bias, ensures public safety, and responsibly empowers developers. We look forward to partnering with your office and key stakeholders to transform this policy into a workable, bipartisan bill.

Works Cited

179. (2020, April 1). *Predictive policing explained*. Brennan Center for Justice.
<https://www.brennancenter.org/our-work/research-reports/predictive-policing-explained>
- Ai Act*. Shaping Europe's digital future. (n.d.).
<https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>
- Arizona becomes first state in U.S. to offer fintech regulatory sandbox*. Arizona Attorney General's Office. (2018, March 22).
<https://www.azag.gov/press-release/arizona-becomes-first-state-us-offer-fintech-regulatory-sandbox>
- Artificial Intelligence in predictive policing issue brief*. NAACP. (2024, February 15).
https://naacp.org/resources/artificial-intelligence-predictive-policing-issue-brief?utm_source=chatgpt.com
- Companies' policy advocacy*. AI Lab Watch. (n.d.). <https://ailabwatch.org/resources/company-advocacy/>
- Google. (n.d.). *Introduction to model evaluation for fairness | vertex AI | google cloud*. Google.
https://cloud.google.com/vertex-ai/docs/evaluation/intro-evaluation-fairness?utm_source=chatgpt.com
- Group, B. C. (2022, January 26). *Utah's new regulatory sandbox*. Consumer Finance Monitor.
<https://www.consumerfinancemonitor.com/2019/06/11/utahs-new-regulatory-sandbox/>
- Guardian News and Media. (2024, December 14). *She didn't get an apartment because of an AI-generated score – and sued to help others avoid the same fate*. The Guardian.
<https://www.theguardian.com/technology/2024/dec/14/saferent-ai-tenant-screening-lawsuit>
- Microsoft responsible AI Impact Assessment Guide. (n.d.-a).
<https://blogs.microsoft.com/wp-content/uploads/prod/sites/5/2022/06/Microsoft-RAI-Impact-Assessment-Guide.pdf>
- Newsom, G. (n.d.). SB-1047-veto-message.PDF.
<https://www.gov.ca.gov/wp-content/uploads/2024/09/SB-1047-Veto-Message.pdf>
- Regulatory sandboxes in Artificial Intelligence | OECD. (n.d.-b).
https://www.oecd.org/en/publications/regulatory-sandboxes-in-artificial-intelligence_8f80a0e6-en.html
- Stevens, C., & Holmes, J. (2023, November 13). *Complying with New York City's bias audit law*. Nixon Peabody LLP.
https://www.nixonpeabody.com/insights/alerts/2023/11/13/complying-with-new-york-city-bias-audit-law?utm_source=chatgpt.com