Date of Hearing: April 22, 2025
Fiscal: Yes

ASSEMBLY COMMITTEE ON PRIVACY AND CONSUMER PROTECTION
Rebecca Bauer-Kahan, Chair
AB 1018 (Bauer-Kahan) – As Amended April 10, 2025

SUBJECT: Automated decision systems

SYNOPSIS

Automated Decision Systems (ADS) typically use artificial intelligence (AI) that produce simplified outputs – such as scores, classifications, or recommendations – to assist or replace human discretionary decisionmaking. Often these decisions have low-stakes; sometimes, however, they can meaningfully impact the public's rights, opportunities, or access to critical resources or services. ADS can process enormous datasets, identify hidden patterns, and make decisions with efficiency and scale that vastly exceeds human capabilities. But relying on ADS to make consequential decisions can be hazardous if the systems are not trained carefully or tested thoroughly: the datasets they are trained on are often unrepresentative or contaminated with bias, the inferences they draw from those datasets are often inscrutable, and these systems can fail to accurately account for the complexity of human behavior. Without human-centered oversight, particularly in consequential contexts such as employment, housing, healthcare, and criminal justice, these impacts can be irreparable.

In 2022, the White House Office of Science and Technology Policy released the Blueprint for an AI Bill of Rights, *which identifies five principles that should "guide the design, use, and deployment of automated systems to protect the American public in the age of artificial intelligence."*[1] *These principles would (1) protect the public from unsafe and ineffective ADS, (2) prevent algorithmic discrimination, (3) protect data privacy, (4) give subjects notice and explanation when ADS are used, and (5) provide for human alternatives, consideration, and fallback. The* Blueprint *calls for rigorous testing, monitoring, and independent evaluation of ADS to ensure that ADS are at least as fair and trustworthy as the humans they are displacing.*

This bill is the author's third attempt to realize that vision. In broad strokes, the bill is similar to its predecessors: it seeks to create a comprehensive transparency regime for developers and deployers of ADS that are used for consequential decisionmaking. The author's goal is to prevent algorithmic discrimination on the basis of protected characteristics and to ensure that ADS can be trusted to perform reliably and accurately. The bill would require deployers of ADS to give subjects of consequential decisions notice, an ability to opt out of the use of the ADS, a chance to correct any personal information used by the ADS to make the decision, and a right to appeal the outcome of the decision. The bill also provides for independent evaluation by third-party auditors – a substantial but controversial provision not found in the prior bills. Enforcement of the bill may be undertaken by the Attorney General, public prosecutors, the Civil Rights Department, or Labor Commissioner, within their respective jurisdictions.

---

[1] The White House, "Blueprint for an AI Bill of Rights," (Oct. 2022), p. 14, https://bidenwhitehouse.archives.gov/ostp/ai-bill-of-rights/ ("Blueprint"). Despite the use of the term "AI" in its title, the Blueprint focuses on ADS.

The bill is co-sponsored by SEIU California and TechEquity and is supported by a broad coalition of civil society, consumer protection, and labor organizations. Supporters argue that the bill establishes common-sense guardrails to prevent inaccurate or discriminatory decisions by ADS – protections that are especially needed after the federal government's retreat from AI regulation.

The bill is opposed by a coalition of industry trade organizations, led by California Chamber of Commerce, who set forth several concerns relating to the bill's breadth, workability, and costs. A number of other businesses and organizations take an oppose-unless-amended position.

If passed by this Committee, this bill will next be heard by the Judiciary Committee.

THIS BILL:

1) Defines key terms, including:

    a) "Automated decision system," or ADS, means a computational process derived from machine learning, statistical modeling, data analytics, or artificial intelligence that issues simplified output, including a score, classification, or recommendation, that is designed or used to assist or replace human discretionary decisionmaking and materially impacts natural persons. The term does not include a spam email filter, firewall, antivirus software, identity and access management tool, calculator, database, dataset, or other compilation of data.

    b) "Consequential decision" means a decision that materially impacts the cost, terms, quality, or accessibility of any of the following to a natural person:

        i) Employment-related decisions, as defined.

        ii) Education and vocational training as they relate to specified categories.

        iii) Housing and lodging as they relate to rentals or short-term housing and lodging, home appraisals, rental subsidies, and publicly supported housing.

        iv) Essential utilities: electricity, heat, water, transportation, and municipal trash and sewage services.

        v) Family planning, adoption services, reproductive services, and assessments related to child protective services.

        vi) Health care and health insurance.

        vii) Financial services.

        viii) The criminal justices system with respect to pretrial release, sentencing, and alternatives to incarceration.

        ix) Legal services, private arbitrations, and mediation.

        x) Elections, as they relate to specified categories.

xi) Access to government benefits or services or assignment of penalties by a government entity.

xii) Places of public accommodation, as specified.

xiii) Insurance.

xiv) Internet and telecommunications access.

c) "Covered ADS" means an ADS that is designed or used to make or facilitate a consequential decision.

d) "Deployer" means a person, partnership, state or local government agency, corporation, or developer that uses a covered ADS to make or facilitate a consequential decision, either directly or by contracting with a third party for that purpose.

e) "Developer" means a person, partnership, state or local government agency, corporation, or deployer that designs, codes, substantially modifies, or otherwise produces an automated decision system that makes or facilitates a consequential decision, either directly or by contracting with a third party for those purposes.

f) "Developer-approved use" means a deployment context in which a developer intends a covered ADS to make or facilitate a consequential decision, and includes any reasonably foreseeable fine tuning of the covered ADS.

g) "Disparate impact" means a differential effect on a group of individuals who share a protected characteristic.

h) "Disparate treatment" means differential treatment of an individual or group of individuals on the basis of a protected characteristic.

i) "Express consent" means affirmative written authorization granted in response to a notice that meets specified requirements.

j) "Fine-tune" means to adjust the model parameters of an ADS through exposure to additional data.

k) "Personal information" has the same meaning as defined in the California Consumer Privacy Act.

l) "Protected characteristic" means a characteristic listed in the Unruh Civil Rights Act.

m) "Substantial modification" means a new version, release, update, or other modification to a covered ADS that materially changes its uses or outputs, but does not include modifications resulting from fine-tuning.

n) "Trade secret" has the same meaning as provided under the Uniform Trade Secrets Act.

2) Requires developers to:

a) Conduct initial performance evaluations on ADS either before they are first deployed or, if they were deployed before January 1, 2026, by January 1, 2027.

b) Conduct subsequent performance evaluations on ADS no more than a year apart, including following any substantial modification or fine-tuning.

c) In conducting a performance evaluation, for each developer-approved use, document:

   i) The expected accuracy and reliability of the covered ADS.

   ii) Whether any disparate treatment is intended to occur, whether such treatment is necessary, and whether alternatives were considered.

   iii) Whether the covered ADS is reasonably likely to result in disparate impacts, whether these impacts are necessary, and whether alternatives were considered.

   iv) Whether measures were taken to mitigate the risk of unanticipated disparate impacts, including any unanticipated disparate impacts reported by deployers.

   v) Any reasonably foreseeable effects of fine tuning, as specified.

d) Contract with an independent third-party auditor to assess the developer's compliance with requirements for performance evaluations. Provide the auditor with necessary documentation, which may be redacted to protect trade secrets. Consider and attempt to incorporate auditor feedback into subsequent versions of the ADS. Make a high-level summary of the feedback publicly available.

e) Disclose to deployers results of the most recent performance evaluation, instructions for using and fine tuning the ADS, an explanation of circumstances in which a deployer's fine tuning of the ADS requires them to assume the responsibilities of a developer, and any technical information needed to comply with the bill.

f) Provide specified information received from audits of performance evaluation to deployers, including discrepancies from anticipated performance, unanticipated disparate impacts, and steps the deployer can take to mitigate discrepancies.

g) Consider and attempt to incorporate feedback from auditors into development of subsequent versions of covered ADS.

h) Ensure documentation provided to deployers is provided in a clear manner in a language commonly used with the developers, and that such documentation is retained for as long as the ADS remains deployed or available to potential deployers plus 10 years.

i) Refrain from representing that an ADS can perform in a manner not substantiated by the results of the most recent performance evaluation, which under the bill constitutes false advertising.

j) Designate at least one employee to oversee compliance with the bill's requirements.

3) As of January 1, 2027, requires deployers to:

a) Before finalizing a consequential decision, provide subjects of the decision:

   i) Written notice with information about the covered ADS, including the personal characteristics it measures or assesses, sources of personal information collected from the subject, structure and format of the outputs, whether a natural person reviews the outputs or outcome before finalizing the consequential decision, and the subject's right to opt out.

   ii) A reasonable opportunity to opt out of the use of the ADS to make or facilitate the consequential decision.

   iii) A deployer need not comply with (i) or (ii) if the subject is having a medical emergency or the deployer is subject to the federal Gramm-Leach-Bliley Act, which governs financial institutions.

b) After finalizing a consequential decision, provide the subjects of the decision:

   i) Another written notice similar to the one described above, within 5 days.

   ii) An opportunity to, within 30 days, correct incorrect personal information or appeal the outcome, as specified. If the exercise of either right would result in a different outcome, the deployer must rectify the outcome within 30 days of making this determination. If not, deployers must inform the subject. Deployers may deny a request to correct information but must provide the subject with an explanation for the basis of the denial and, for requests to correct the information, allow them a reasonable opportunity to request deletion of their personal information.

c) Minimize their collection, use, retention, and sharing of personal information from subjects of consequential decisions.

d) If the ADS impacts more than 5,999 people in a three-year period, contract with a third-party independent auditor to conduct an impact assessment on the covered ADS before January 1, 2030 and every three years thereafter.

e) Assume the responsibilities of the developer if the deployer uses a covered ADS that impacts 5,999 or more people in a given three-year period and does not receive documentation relating to the developer-approved uses and performance evaluations for the covered ADS, or substantially modifies the ADS and uses it to impact 5,999 people in a three year period or otherwise makes the modified system available to potential deployers.

f) Retain specified documentation for as long as the ADS is deployed plus 10 years.

g) Designate at least one employee to oversee compliance with the bill.

h) Comply with any additional requirements related to covered ADS promulgated by the California Privacy Protection Agency in duly adopted regulations.

4) Requires auditors conducting impact assessments every three years on behalf of deployers who use covered ADS that impact more than 5,999 people in that period to:

a) Document: each developer-approved use of the covered ADS the deployer utilized; differences in observed and expected accuracy and reliability; whether disparate impacts resulted from the use of the ADS; whether the deployer used the ADS outside the scope of a developer-approved use; and whether the deployer assumed the responsibilities of a developer.

b) Provide the results of the impact assessment to the developer and deployer.

c) Make a high-level summary of the results of the impact assessment publicly available.

5) Enables the Attorney General to obtain an unredacted performance evaluation or impact assessment, and to share that documentation with other enforcement entities as necessary for enforcement purposes. Such documentation is exempt from the California Public Records Act.

6) Enables public prosecutors, the Civil Rights Department, and Labor Commissioner (with respect to employment-related violations) to bring a civil action against a developer or deployer to bring an action for injunctive or declaratory relief, reasonable attorney's fees and costs, and a civil penalty of up to $25,000 per violation.

7) Provides that a developer or deployer who contracts with a third party to comply with duties required under the bill, other than duties relating to auditing, is liable for the third party's failure to comply with those duties.

8) Does not:

a) Apply to ADS that solely serve a cybersecurity function or operate aircraft in the national airspace.

b) Apply to use of a consumer credit score.

c) Limit rights, remedies, and penalties under any other law.

d) Authorize use of an ADS, or disparate impacts and treatments, otherwise limited, restricted, or prohibited under any other law.

9) Provides that in an action alleging a violation of the Unruh Civil Rights Act or FEHA in which an ADS is alleged to have committed or facilitated the violation, the defendant's compliance with this bill is relevant to, but not conclusive of, the claim.

EXISTING LAW:

1) Establishes the Civil Rights Department, and sets forth its statutory functions, duties, and powers. (Gov. Code § 12930.)

2) Establishes the Fair Employment and Housing Act. (Gov. Code § 12900 *et seq*.)

3) Establishes the Unruh Civil Rights Act, which provides that all persons within the jurisdiction of this state are free and equal, and no matter what their sex, race, color, religion, ancestry, national origin, disability, medical condition, genetic information, marital status, sexual orientation, citizenship, primary language, or immigration status are entitled to the full

and equal accommodations, advantages, facilities, privileges, or services in all business establishments of every kind whatsoever. (Civil Code § 51(b).)

4) Defines "trade secret" under the Uniform Trade Secrets Act as information, including a formula, pattern, compilation, program, device, method, technique, or process, that both:

   a) Derives independent economic value, actual or potential, from not being generally known to the public or to other persons who can obtain economic value from its disclosure or use; and

   b) Is the subject of efforts that are reasonable under the circumstances to maintain its secrecy. (Civ. Code § 3426.1(d).)

5) Establishes the California Privacy Protection Agency (Privacy Agency) and vests it with full administrative power, authority, and jurisdiction to implement and enforce the California Consumer Privacy Act (CPPA) of 2018. (Civ. Code § 1798.199.10.)

6) Requires the California Privacy Protection Agency to issue regulations governing access and opt-out rights with respect to businesses' use of automated decisionmaking technology. (Civ. Code § 1798.185.)

7) Establishes prohibitions on false and deceptive advertising. (Bus. & Prof. Code § 17500 et seq.)

COMMENTS:

1) **Author's statement**. According to the author:

> The use of automated decision systems (ADS) has become prevalent among Californian's daily lives and used within various sectors – including, housing, employment, and even in criminal justice sentencing and probation decisions. However, the algorithms that power ADS are often vulnerable to issues such as unrepresentative data, faulty classifications, and flawed design. These shortcomings can result in biased, discriminatory, or unfair outcomes. Rather than solving systemic problems, poorly designed ADS can worsen the very harms they aim to address—ultimately hurting the people they are meant to help. AB 1018 provides the necessary guardrails by regulating the development and deployment of an ADS used to make consequential decisions. Specifically, it requires developers to conduct an initial performance evaluation of an ADS by January 1, 2027. Additionally, the current role that an ADS plays in a consequential decision is hidden from consumers. Fundamentally, consumers should have the right to be well informed of how these ADS are used in life altering decisions. AB 1018 calls for transparency on the usage of ADS by requiring notice to consumers by deployers before and after a consequential decision. It is crucial that we take the necessary steps to ensure the technology is used responsibly and can be trusted. Well-intentioned but flawed technology is a matter of state concern. Guardrails and accountability are needed to ensure that technology does not further marginalize communities or broaden inequities.

2) **Automated Decision Systems.** As a general matter, ADS use AI to make life-impacting decisions. AI refers to the mimicking of human intelligence by artificial systems such as computers. AB 2885 (Bauer-Kahan, Stats. 2024, Ch. 843) defined the term as "an engineered or

machine-based system that varies in its level of autonomy and that can, for explicit or implicit objectives, infer from the input it receives how to generate outputs that can influence physical or virtual environments." AI uses algorithms – sets of rules – to transform inputs into outputs. Inputs and outputs can be anything a computer can process: numbers, text, audio, video, or movement. AI is not fundamentally different from other computer functions; unlike other computer functions, however, AI is able to accomplish tasks that are normally performed by humans.

Most modern AI tools are created through a process known as "machine learning." Whereas traditional programming is manually written by a programmer and can only follow pre-defined rules, with machine learning algorithms, programmers do not determine the rules the algorithm will follow. Instead, the AI is allowed to automatically explore the structure of data and infer the relationship between the data and desired outputs.[2] Machine-learning algorithms are often criticized for being "black boxes" that generate predictions and outcomes that cannot be clearly explained.[3] "It's often observed in the field that the most powerful models are on the whole the least intelligible, and the most intelligible are among the least accurate."[4]

The process of exposing a naïve AI to data is known as "training." The algorithm that an AI develops during training is known as its "model." Training is the secret sauce of machine learning. At its core, training is an optimization problem wherein a model attempts to identify specific parameters – "weights" – that minimize the difference between predicted outcomes and actual outcomes. How an input is transformed into an output depends on the specific algorithm that is developed by a model. Once trained, AI can be used to process new, never-before-seen data.

AI that are trained on small, specific datasets in order to make recommendations and predictions are sometimes referred to as "predictive AI." This differentiates them from generative AI (GenAI,) which are trained on massive datasets in order to produce detailed text, images, audio, and video. When ChatGPT generates text in clear, concise paragraphs, it uses GenAI that is trained on the written contents of the internet.[5] When Netflix suggests content to a viewer, its recommendation is produced by predictive AI that is trained on the viewing habits of Netflix users.[6]

ADS typically use predictive AI to generate outputs that influence real-world outcomes.[7] Existing law defines an ADS to include "a computational process derived from machine learning, statistical modeling, data analytics, or artificial intelligence that issues simplified output, including a score, classification, or recommendation, that is used to assist or replace human discretionary decisionmaking and materially impacts natural persons."[8]

---

[2] IBM, *What is machine learning?* www.ibm.com/topics/machine-learning.
[3] Neil Savage, "Breaking into the black box of artificial intelligence," *Nature* (March 29, 2022).
[4] Brian Christian, "The Alignment Problem: Machine Learning and Human Values" (Norton 2020, 1st ed.), p. 85.
[5] OpenAI, *How ChatGPT and Our Language Models Are Developed*,https://help.openai.com/en/articles/7842364-how-chatgpt-and-our-foundation-models-are-developed.
[6] Netflix, *How Netflix's Recommendations System Works*, https://help.netflix.com/en/node/100639.
[7] Hany Farid, "Artificial Intelligence: A Primer for Legal Practitioners" *Artificial Intelligence: Legal Issues, Policy, and Practical Strategies* (American Bar Association, 2024), p. 73.
[8] Gov. Code § 11546.45.5(a)(1).

When the stakes of these outcomes are high, such as predicting a person's risk of defaulting on a loan or committing a crime, the use of an ADS can cut both ways: these systems can create efficiencies, ensure uniformity, and detect intricate patterns among massive datasets, leading to incredibly beneficial applications and breakthroughs.[9] On the other hand, use of ADS can also raise significant questions of fairness, accountability, and transparency. Government Code section 11546.45.5(a)(4) defines "high-risk" ADS as those that are "used to assist or replace human discretionary decisions that have a legal or similarly significant effect, including decisions that materially impact access to, or approval for, housing or accommodations, education, employment, credit, health care, and criminal justice."

This bill incorporates the existing definition of ADS and applies to any such system used to make or facilitate a "consequential decision," defined under the bill as a decision that materially impacts the cost, terms, quality, or accessibility of several enumerated categories with respect to a natural person.

3) **Algorithmic discrimination.** There is a well-known saying in computer science: "garbage in, garbage out." The performance of an ADS is directly impacted by the quality, quantity, and relevance of the data used to train it.[10] If the data used to train the ADS is biased, the tool's outputs will be similarly biased. Over the past thirty years, several industries have been forced to contend with this fact as they have attempted to introduce ADS into their workflows. Specific examples follow.

Child welfare: In 2016, Allegheny County, Pennsylvania adopted a family screening tool to predict which families should be investigated by social workers for possible removal of maltreated children. The tool was trained only on data from poorer families who used public services such as Medicaid. Because its training dataset lacked examples of wealthier families, the tool disproportionately targeted poorer families.[11]

Credit and loan approval. Financial tools that utilize ADS are similarly susceptible to bias and discrimination. An investigation by The Markup (and co-published by the Associated Press) revealed that in 2019, lenders were more likely to deny home loans to people of color; in particular, lenders were 40 percent more likely to turn down Latino applicants for loans, 50 percent more likely to deny Asian/Pacific Islander applicants, and 70 percent more likely to deny Native American applicants than similar white applicants. Lenders were 80 percent more likely to reject Black applicants than similar white applicants. In every case, the prospective borrowers of color looked almost exactly the same on paper as the white applicants, except for their race. According to the report, mortgage-approval algorithms play a major role in perpetuating these inequities.[12]

Education. Professor Broussard's article in the *New York Times* details how "the International Baccalaureate—a global program that awards prestigious diplomas to high school students—

---

[9] See e.g. Santariano & Metz, "Using A.I. to Detect Breast Cancer That Doctors Miss," *New York Times* (Mar. 5, 2023), https://www.nytimes.com/2023/03/05/technology/artificial-intelligence-breast-cancer-detection.html.
[10] Rohit Sehgal, "AI Needs Data More Than Data Needs AI," *Forbes* (Oct. 5, 2023), https://www.forbes.com/sites/forbestechcouncil/2023/10/05/ai-needs-data-more-than-data-needs-ai/.
[11] Arvind Narayanan and Sayash Kapoor, *AI Snake Oil: What Artificial Intelligence Can Do, Can't Do, and How to Tell the Difference* (1st ed. 2024), pp. 52-53.
[12] Emmanuel Martinez and Lauren Kirchner, "The Secret Bias Hidden in Mortgage-Approval Algorithms," *Markup* (Aug. 25, 2021), https://themarkup.org/denied/2021/08/25/the-secret-bias-hidden-in-mortgage-approval-algorithms.

canceled its usual in-person exams because of the [COVID-19] pandemic" and instead "used an algorithm to 'predict' student grades based on an array of student information, including teacher-estimated grades and past performance by students in each school."[13] Tens of thousands of students, surprised to find out they failed, protested the results. "High-achieving, low-income students were hit particularly hard: many took the exams expecting to earn college credit with their scores and save thousands of dollars on tuition."[14]

Employment. It is no secret that people of various races, genders, and cultures are not distributed equally throughout the workforce. An ADS that is trained on historical data to make hiring decisions will be predisposed to maintain the ratios it is trained on; as described by Aditya Malik, the Founder and CEO of Valuematrix.ai:

> Generative AI, for all its grandeur, has the potential to perpetuate latent biases inherited from human creators. A disconcerting echo of historical prejudices may inadvertently seep into the algorithms. Imagine a scenario where previous senior managers, driven by biases of gender, age, faith or race, rejected candidates for misguided reasons. The AI, if not vigilantly curated, might misconstrue these patterns as indicators of incompetence, thus exacerbating the exclusion of qualified candidates from underrepresented backgrounds.[15]

This was notoriously experienced by Amazon, who considered automating their hiring practices in the early 2010s. They opted against this approach in 2015 when they realized that their ADS-enabled system was not rating candidates in a gender-neutral way. In fact, their system was excluding women from the pool of acceptable candidates because it had been trained to vet applicants by observing patterns in resumes submitted to the company over a 10-year period. Most came from men, a reflection of inequities across the tech industry.[16]

ADS are being employed in the workforce as well. For example, "[a] company installed AI-powered cameras in its delivery vans in order to evaluate the road safety habits of its drivers, but the system incorrectly penalized drivers when other cars cut them off or when other events beyond their control took place on the road. As a result, drivers were incorrectly ineligible to receive a bonus."[17]

Healthcare. When ADS are deployed in healthcare, biased historical data can lead to patients being recommended substandard care on the basis of their race or ethnicity. In 2007, an ADS was developed to help doctors estimate whether it was safe for people who had delivered previous children through cesarean section to deliver subsequent children vaginally – a risky procedure. The ADS considered various health relevant factors as it made its decision, such as the woman's age, her reason for the previous cesarean, and how long ago the cesarean had been performed. However, a 2017 study found that the ADS was biased; it predicted Black and Latino

---

[13] "When Algorithms Give Real Students Imaginary Grades," *New York Times* (Sept. 8, 2020), https://www.nytimes.com/2020/09/08/opinion/international-baccalaureate-algorithm-grades.html.
[14] *Id.*
[15] Aditya Malik, "AI Bias In Recruitment: Ethical Implications And Transparency," *Forbes* (Sep. 25, 2023), https://www.forbes.com/sites/forbestechcouncil/2023/09/25/ai-bias-in-recruitment-ethical-implications-and-transparency/.
[16] Jeffrey Dastin, "Amazon scraps secret AI recruiting tool that showed bias against women," *Reuters* (Oct. 9, 2018), https://www.reuters.com/article/amazoncom-jobs-automation/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSL2N1VB1FQ/.
[17] *Blueprint, supra,* p. 17.

people were less likely to have a successful vaginal birth after a cesarean than non-Hispanic white women. As a result, doctors performed more cesareans on Black and Latino people than on white people.[18] These discrepancies perpetuate historical biases – Black Americans, for example, have historically received a lower standard of healthcare than their white counterparts.[19]

Similarly, a 2019 study found strong racial bias in a system used to identify patients with a high risk of adverse health outcomes. The ADS assigned Black patients lower likelihoods than equally at-risk White patients. The authors found that this happened because the ADS was designed to predict healthcare costs instead of needs. Because the healthcare system has historically spent less on care for Black patients than White patients for the same health conditions, the ADS was, in essence, issuing a prediction that mirrored and perpetuated past discrimination.[20]

Housing. A recent *ProPublica* article found that tenant screening companies compile information on renters beyond credit reports – including criminal background, evictions filings, medical debt, and student loans – and then use algorithms that "try to predict how risky it is to rent to a potential tenant based on characteristics they share with other tenants" and "assign applicants scores or provide landlords a yes-or-no recommendation."[21] In 2023, concerns about these tools led several Attorneys General – California's Rob Bonta included – to write in a comment letter to the Federal Trade Commission and the Consumer Financial Protection Bureau:

> Landlords and tenant screening companies increasingly rely on problematic screening algorithms that combine various information to generate a single score or result indicating how "safe" it would be to rent to a prospective tenant. These algorithms lack transparency and can be inaccurate, and can have a discriminatory impact on underserved communities. The attorneys general recommend, among other things, requiring that tenant screening companies disclose their reliance on algorithms and screen algorithm models for bias against protected classes and prohibiting the use of certain types of records in screening reports.[22]

Sentencing and bail decisions. ADS are frequently used to inform sentencing and bail decisions. These tools are trained using historical data, and the predictions they make can therefore reflect historical bias. A 2016 ProPublica study dove into the use of one such tool – COMPAS – in Broward County, Florida.[23] The study determined that Black defendants were far more likely

---

[18] Caleb J Colón-Rodríguez, "Shedding Light on Healthcare Algorithmic and Artificial Intelligence Bias," *U.S. Department of Health & Human Services Office of Minority Health,* (Jul. 12, 2023), https://pmc.ncbi.nlm.nih.gov/articles/PMC6875681/.

[19] California Task Force to Study and Develop Reparation Proposals for African American, "Final Report," p. 461, https://oag.ca.gov/system/files/media/full-ca-reparations.pdf.

[20] Obermeyer et al, "Dissecting racial bias in an algorithm used to manage the health of populations," *Science* 2019., 366(6464):447–453.

[21] Erin Smith and Helen Vogell, "How Your Shadow Credit Score Could Decide Whether You Get an Apartment" *ProPublica* (Mar. 29, 2022), https://www.propublica.org/article/how-your-shadow-credit-score-could-decide-whether-you-get-an-apartment. While the mere compilation and provision of background information is not subject to this bill, processing that information through an algorithm to produce a simplified output such as a score or yes/no recommendation for a prospective tenant – thereby outsourcing human discretionary decisionmaking – generally falls under this bill.

[22] "Attorney General Bonta Submits Comment Letter Recommending Reforms to the Tenant Screening Process" (May 31, 2023), https://oag.ca.gov/news/press-releases/attorney-general-bonta-submits-comment-letter-recommending-reforms-tenant#:~:text=OAKLAND.

[23] Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, "Machine Bias," *Propublica*, (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

than white defendants to be incorrectly judged to be at a higher risk of recidivism, while white defendants were more likely than Black defendants to be incorrectly flagged as low risk.

The for-profit company that developed this tool, Northpointe, does not publicly disclose the calculations used to arrive at defendants' risk scores, so it is not possible for either defendants or the public to see what might be driving the disparity. These discrepancies mirror historical injustices perpetuated against Black Americans by California's criminal justice system.[24] The use of a proprietary algorithm by government actors also raises significant due process questions, as it becomes difficult for individuals to understand, let alone challenge, arbitrary government deprivations of life, liberty, or property.[25]

Structural effects of discriminatory ADS: ADS becoming more pervasive across institutions raises the risk of a feedback loop that compounds and reinforces discrimination. As Borocas et al. write in *Fairness and Machine Learning*, "predictive systems have the effect of transferring advantages from one phase of life to the next, and one generation to the next." This can compound and amplify injustices: "individuals are subject to a series of decisions over the course of their lives, and the effects of these decisions both accumulate and compound over time. When a person receives (or is denied) one opportunity, they are likely to appear more (or less) qualified at their next encounter with a predictive system."[26]

4) **Unsafe or ineffective systems.** In addition to discriminatory outcomes, some ADS are unsafe or simply ineffective regardless of who the subjects of the prediction are. A few common types of flawed ADS are described below.

Spurious correlations. Accurate predictions may nevertheless lead to bad decisions. In one example, a hospital trained AI models on a dataset of 15,000 pneumonia patients in order to develop a model that could identify which pneumonia patients were at the greatest risk in order to triage new patients. During testing, it was discovered that one of the most accurate models recommended outpatient status for asthmatics – a life-threateningly dangerous error based on an accurate statistical correlation: asthmatics are less likely to die from pneumonia than the general population precisely because asthma is such a serious risk factor that asthmatics automatically get elevated care.[27]

In other cases, the correlations that machine-learning ADS may rely on have little to do with the attributes they purport to measure. Some hiring tools record videos of applicants responding to pre-recorded questions and attempt to assess their fitness for a job based on their facial expression and intonation. Such tools can be gamed by making simple changes to the subject's appearance, such as wearing glasses, or the background of the room, such as adding more books on the bookshelf – leading to increased scores. Journalist Hilke Schellmann found she was able

---

[24] California Task Force to Study and Develop Reparation Proposals for African American, "Final Report," *supra,* p. 420.
[25] See *Mathews v. Eldridge* (1976) 424 U.S. 319, 333; Daniel Keats Citron, Technological Due Process (2007) 85 Wash. U. L. Rev. 1249, 1249 https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1012360
[26] Boracas et al., *Fairness and Machine Learning: Limitations and Opportunities* (2023) pp. 213-214.
[27] Brian Christian, "The Alignment Problem: Machine Learning and Human Values" (Norton 2020, First Ed.), pp. 82-84.

to obtain consistently high scores on one hiring ADS despite responding by reading an irrelevant Wikipedia entry in German.[28]

Irrelevant datasets. "AI reflects its training data. It learns patterns about the people who make up the data, and the decisions made by AI reflect these patterns. But when the decision subjects come from a population with different characteristics than those in the training data, the model's decisions are likely to be wrong."[29] For instance, the Ohio Risk Assessment System was trained on data from just 452 defendants from Ohio, but has been deployed in several other states, despite its small and unrepresentative dataset.[30]

Snake oil. Some tools, although marketed as automating precision, are simply are not effective. In 2022, Toronto used an ADS to predict when high bacteria levels made it unsafe to swim at public beaches. Although the developer claimed the tool as 90 percent accurate, it fared far worse: "on 64 percent of the days when the water was unsafe, beaches remained open based on incorrect assessments." Yet officials never overrode the recommendations produced by the tool.[31] Similarly, in 2017, a sepsis prediction tool that was deployed in hundreds of hospitals across the U.S. Despite having high accuracy when it was internally tested, a 2021 study found the tool missed two-third of the sepsis cases and led to a high rate of false alerts.[32]

As Princeton researchers Arvind Narayanan and Sayash Kapoor put it bluntly in *AI Snake Oil:* "In contrast to generative AI, predictive AI often does not work at all."[33] As described below, such tools are especially questionable when used to forecast individual human behavior.

5) **"It's tough to make predictions, especially about the future."**[34] Some phenomena, such as the movement of the planets, are easy to predict because they follow the laws of physics. Others, such as the weather, are chaotic systems that are harder to predict but can be modeled to make short-term predictions. Often, human behavior is even less predictable. Although some aggregate social patterns, such as traffic flow, are relatively consistent, how individuals will exercise their own agency cannot be predicted with certainty. How do we know if an individual will, for example, pay a debt? Even if they are fully committed to honoring the debt, unforeseen circumstances – illness or job loss – may prevent them from following through.

When it comes to predicting human behavior, ADS often fail to outperform simple statistical models. For instance, the Fragile Families challenge used over 10,000 data points about 4,000 children in 20 different U.S. cities to see how well AI could predict outcomes for specific children, such as their GPA. Complex AI models trained on all of this data fared no better than an analysis of a few key variables, such as the mother's educational attainment.[35] Similarly,

---

[28] *See* Hilke Schellmann, "The Algorithm: How AI Decides Who Gets Hired, Monitored, Promoted, and Fired and Why We Need to Fight Back Now" (1st ed. 2024).
[29] *Snake Oil, supra,* at p. 73.
[30] *Id.* at p. 51.
[31] *Id.* at p. 50.
[32] Wong et al, "External valiation of a widely implemented proprietary sepsis prediction model in hospitalized patients" JAMA Int. Med. 181 (Aug. 2021), 1065–1070, https://doi.org/10.1001/jamainternmed.2021.2626.
[33] *Snake Oil, supra,* at p. 9.
[34] Quote from Yogi Berra.
[35] *Snake Oil, supra,* at p. 73.

COMPAS, which used 137 data points to assess recidivism rates, fared no better than two data points about an individual – their age and number of prior offenses.[36]

A recently released study compiled a list of 47 applications of ADS that use machine learning to predict the future behavior or outcomes for individuals in eight domains: criminal justice, healthcare, welfare, finance, education, workplace, marketing, and recommender systems. The study concluded that such tools frequently fall well short of their purported benefits. The authors argue that developers and deployers of such systems should have the burden of demonstrating that their tools are not harmful.[37] As Narayanan and Kapoor write: "Accurately predicting people's social behavior is not a solvable technology problem and determining people's life chance on the basis of inherently faulty predictions will always be morally problematic."[38]

Nevertheless, some AI companies market their products as capable of predicting future human behavior. For example, in 2019 RealPage announced its "AI Screening" tool as follows:

> Traditional screening models use credit score, rent-to-income, debt-to-income and generic financial data to determine renter risk. While these factors broadly measure an applicant's **capability to pay financial obligations**, including rent, RealPage developed industry-specific insights to determine the **willingness to pay rent**. Together, analyzing an applicant's capability and willingness to pay rent is a superior risk assessment model to predict a renter's financial performance.[39]

In 2023, the Federal Trade Commission warned companies that "we're not yet living in the realm of science fiction, where computers can generally make trustworthy predictions of human behavior. Your performance claims would be deceptive if they lack scientific support or if they apply only to certain types of users or under certain conditions."[40]

6) **This bill creates a comprehensive transparency regime for ADS used in life-impacting decisions.** Under the bill, an ADS is a computational process that produces "simplified outputs" – such as a score, recommendation, or classification – that are used to assist or replace human discretionary decisionmaking. The bill applies to a subset of ADS – "covered ADS" – that are used to make or facilitate a consequential decision. A consequential decision, in turn, is one that materially impacts the cost, terms, quality, or accessibility of any of several enumerated categories.

With respect to covered ADS, developers must conduct at least yearly performance evaluations that disclose the following:

- The expected accuracy and reliability of the covered ADS.
- Whether any disparate treatment is intended to occur, whether such treatment is necessary, and whether alternatives were considered.

---

[36] *Id.* at p. 80.
[37] Angelina Wang et al. 2023. Against predictive optimization: On the legitimacy of decision-making algorithms that optimize predictive accuracy. In Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (Chicago, IL, USA: ACM, 2023), 626–26.
[38] Snake Oil, 15.
[39] "RealPage Release AI Screening," (Jun. 26, 2019), https://www.realpage.com/news/realpage-releases-ai-screening/. Emphasis in original.
[40] *Snake Oil, supra,* 25.

- Whether the covered ADS is reasonably likely to result in disparate impacts, whether these impacts are necessary, and whether alternatives were considered.
- Whether measures were taken to mitigate the risk of unanticipated disparate impacts, including any unanticipated disparate impacts reported by deployers.
- Any reasonably foreseeable effects of fine tuning, as specified.

These evaluations are limited to "developer-approved uses" – the purposes for which the developer intends the ADS to be used. The developer is prohibited from representing that the ADS can perform in a manner not substantiated by the most recent performance evaluation conducted on the system. In subsequent performance evaluations, developers must consider feedback about the ADS's performance from deployers. The developer must contract with independent third-party auditors to evaluate the developer's compliance with their obligations related to performance evaluations. The developer must make a high-level summary of the auditor's feedback publicly available. Developers must designate at least one employee to oversee compliance.

Deployers are required to provide subjects with pre-and post-decision notice that includes specified information about the covered ADS and the subject's rights, as well as a reasonable opportunity to opt out of the use of the ADS. After the decision is made, subjects have 30 days to correct incorrect personal information used by the ADS or appeal the outcome. If the ADS affects more than 5,999 in a three-year period, deployers must, at the end of that period, contract with an independent third-party auditor to conduct an impact assessment. Unlike performance evaluations undertaken by developers, which are forward-looking evaluations of the model's overall anticipated performance, impact assessments are a backwards-looking assessment of the ADS's actual impact in the real world. Auditors performing impact assessments must document:

- Each developer-approved use of the covered ADS the deployer utilized.
- Differences in observed and expected accuracy and reliability.
- Whether disparate impacts resulted from the use of the ADS.
- Whether the deployer used the ADS outside the scope of a developer-approved use.
- Whether the deployer assumed the responsibilities of a developer by using the ADS, including by using the ADS outside the scope of developer-approved uses on more than 5,999 people over the course of the three-year period.

These impact assessments are provided to the deployer and the developer, who must take the results into account in future performance evaluations. A high-level summary of these impact assessments must be made publicly available. In certain circumstances, deployers that use ADS in specified ways assume the duties of a developer. Deployers must also designate at least one employee to oversee compliance.

To protect personal information, deployers would be required to minimize their collection, use, retention, and sharing of personal information used for consequential decisions. The bill also provides that it does not require the collection of information that is not otherwise collected in the ordinary course of business. Auditors of deployed systems may provide personal information of subjects to the developer of the system only if the subject first grants express consent.

The AG may obtain un-redacted performance evaluations or impact assessments, and share these with other enforcement entities as necessary. This documentation is exempted from Public Records Act requests. Public prosecutors, the Civil Rights Department, and Labor Commissioner

may bring a civil action against a developer or deployer to bring an action for injunctive or declaratory relief, reasonable attorney's fees and costs, and a civil penalty of up to $25,000 per violation. Developers and deployers who contract with third parties to perform duties under the bill are liable for the third parties' failure to perform those duties. Auditors are also subject to liability under the bill.

The bill expressly provides that it does not override other rights, remedies or penalties under other provisions of law. The bill does not apply to ADS that solely serve cybersecurity functions or operate aircraft the national airspace. The bill also provides that compliance with the bill's provisions is relevant to, but not conclusive of, anti-discrimination claims under the Unruh Civil Rights Act or FEHA.

7) **Implementing the *Blueprint for an AI Bill of Rights*.** The provisions of this bill are derived from the principles outlined in the *Blueprint for an AI Bill of Rights,* issued by the White House Office of Science and Technology Policy on October 4, 2022, and adopted by the Legislature in SCR 17 (Dodd, 2023). The *Blueprint* identifies five principles that should "guide the design, use, and deployment of automated systems to protect the American public in the age of artificial intelligence."[41] The applicable principles, and the manner in which the bill implements them, are as follows:

Safe and effective systems. The *Blueprint* summarizes this principle as meaning that "you should be protected from unsafe or ineffective systems."[42] This entails ongoing safety testing, including independent evaluation, and risk mitigation.[43] The bill would implement this principle by requiring developers to conduct forward-looking performance evaluations, subject to independent third-party audits. The bill would also require certain deployers of systems that affect 6,000 or more people to contract every three years with independent third-party auditors to conduct backward-looking impact assessments. Information would flow between developers and deployers to help identify and mitigate bias or ineffective or unsafe systems. Once completed, this documentation would be available to the Attorney General to review and, where necessary, initiate enforcement action. To ensure transparency and accountability, high-level summaries of audits must be made public.

Algorithmic discrimination protections. The *Blueprint* provides that "you should not face discrimination by algorithms[,] and systems should be used and designed in an equitable way."[44] Accordingly, "Automated systems should be tested using a broad set of measures to assess whether the system components, both in pre-deployment testing and in-context deployment, produce disparities."[45] The bill would implement this principle by requiring the testing and independent evaluation described above to determine whether the system is anticipated or found to have disparate impacts on those with protected characteristics. Developers must also disclose whether they took affirmative steps to mitigate potential disparate impacts. Because this creates a record of known potential discriminatory impacts that can be used in antidiscrimination claims, this creates a strong incentive to avoid releasing or using potentially discriminatory systems.

---

[41] *Blueprint, supra,* p. 14.
[42] *Id.* at p. 15.
[43] *Ibid.*
[44] *Id.* at p. 23.
[45] *Id.* at p. 27.

Data privacy. The *Blueprint* identifies data privacy as a "foundational and cross-cutting principle required for achieving all others in this framework."[46] This bill implements this principle by requiring deployers to minimize their collection, use, retention, and sharing of personal information from subjects of consequential decisions. Auditors of deployed systems may provide personal information of subjects to the developer of the system only if the subject first grants express consent. Finally, the bill expressly provides that it does not mandate collection of information beyond that gathered in the ordinary course of business.

Notice and explanation. The *Blueprint* summarizes this principle as meaning that "you should know that an automated system is being used and understand how and why it contributes to outcomes that impact you."[47] The bill would implement this principle by requiring deployers, before finalizing the consequential decision, to provide notice to potential subjects with detailed information about the ADS. Post-decision notice must also be provided within 5 days. This notice gives subjects of consequential decisions notice and the ability to decide whether to exercise additional rights, discussed below.

Human alternatives, consideration, and fallback. The *Blueprint* summarizes this principle as meaning that "[y]ou should be able to opt out, where appropriate, and have access to a person who can quickly consider and remedy problems you encounter."[48] The bill would implement this principle by enabling subjects to opt out of the use of the covered ADS, correct their personal information used by the ADS, and appeal the decision made by the ADS.

The premises of this bill are, first, that Californians ought to be much more informed about the ADS being used to make consequential decisions in their lives, and second, that developers and deployers of ADS ought to be much more conscious and deliberate about what the impacts of their tools might be. In short, California ought to get a handle on ADS while it still can before they become yet another phenomenon – like social media, greenhouse gases, guns, and urban sprawl – that could have been dealt with deliberately and intelligently in the past, but was instead allowed to expand without any constraints past the point of being manageable. These protections are especially important in light of the federal government's recent dismantling of programs that promote trustworthy AI as it has simultaneously demolished efforts to promote fairness and equality.

8) **Comparison to AB 2930.** This bill is substantially similar to the final version of AB 2930,[49] which was moved to the inactive file in the Senate. Key differences follow.

No "substantial factor." AB 2930 applied to ADS that are a "substantial factor" in a consequential decision. "Substantial factor" was defined as "an element of a decisionmaking process that is capable of altering the outcome of the process." This bill more straightforwardly applies to ADS that "make or facilitate" a consequential decision.

Shift from "algorithmic discrimination" to "disparate treatment" or "disparate impact." Under AB 2930, developers and deployers were required to test for "algorithmic discrimination," defined as "unlawful discrimination." Rather than require businesses to formulate and disclose

---

[46] *Id.* at p. 31.
[47] *Id.* at p. 40.
[48] *Id.* at p. 46.
[49] That version was, however, specific to employment, whereas prior versions applied to a broader set of contexts.

what amounts to a legal opinion about the risk of unlawful discrimination, this bill would instead focus on factual disclosures: whether the system is intended to treat protected classes differently, and whether testing shows that it will have a disparate impact on such classes. If developers and deployers use the system despite awareness of substantial and unjustified disparate impacts, they may be subject to a lawsuit on this basis under existing anti-discrimination laws. Certain federal and state laws allow for disparate impact claims; an exception is the Unruh Act, which applies to intentional discrimination only. Nevertheless, in Unruh Act cases, disparate impact "'may be probative of intentional discrimination.'"[50] The bill does not change how anti-discrimination laws work; rather, it seeks to complement these laws through a process that creates a documentary record to help overcome the "black box" problem with many ADS.

"Developer-approved use." This bill introduces the idea of "developer-approved uses" – a deployment context in which the developer intends an ADS to make or facilitate a consequential decision. Developer approved uses are important to the scope of the bill: they define the testing context for performance evaluations. They also limit how the developer may advertise the ADS's capabilities. Finally, if a deployer uses the ADS in a manner that is not consistent with the developer-approved uses, the deployer may assume the developer's responsibilities, ensuring that the developer is not on the hook for unintended uses of the ADS.

Audits. Several AI governance frameworks recommend third-party evaluation of AI tools.[51] The *Blueprint* states calls for third-party audits and states "entities should allow independent evaluation of potential algorithmic discrimination caused by automated systems they use or oversee."[52] As discussed in this committee's recent analysis of AB 1405 (Bauer-Kahan), which would create an enrollment process for AI auditors, auditing is a developing industry that promises to expand more jurisdictions require audits.[53] While AB 2930 did not address audits, this bill effectively requires two types. The first, applicable to developers, is a regulatory compliance audit to ensure they are properly carrying out their duties in conducting performance evaluations. The second, applicable to deployers of broadly used ADS, is an impact assessment that retrospectively evaluates the specific impact of the ADS in the deployment context.

Governance program. Whereas AB 2930 had a prescriptive oversight program for developers and deployers to oversee compliance with the bill's obligations, this bill simply requires that developers and deployers designate an employee for this purpose and requires those employees to promptly review credible compliance issues.

Enforcement. The final version of AB 2930 allowed for administrative and civil enforcement by the Civil Rights Department, including a $25,000 civil penalty. Prior versions of the bill gave public prosecutors enforcement authority. This bill gives publics prosecutors as well as the Labor Commissioner civil enforcement. The civil penalty is *up to* $25,000. The bill does not provide for administrative enforcement.

---

[50] *Liapes v. Facebook, Inc.* (2023) 95 Cal.App.5th 910, 926.
[51] See Selinger et al, "AI Audits: Who, When, How...Or Even If?" (Sep. 2023), in *Collaborative Intelligence: How Humans and AI are Transforming our World,* MIT Press (Forthcoming), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4568208.
[52] *Blueprint, supra,* p. 28; see also p. 19 (third-party auditors to demonstrate safety and effectiveness of system).
[53] E. Mulvaney, "NYC Targets Artificial Intelligence Bias in Hiring Under New Law," Bloomberg Law, 2021, https://news.bloomberglaw.com/daily-labor-report/nyc-tar gets-artificial-intelligence-bias-in-hiring-under-new-law.

9) **Related efforts.** A number of jurisdictions are taking action to regulate ADS. For instance, the European Union AI Act provides guardrails for what it deems "high-risk AI systems" to ensure transparency and fairness. Here in the United States, Senator Markey recently introduced his AI Civil Rights Act, which would impose stringent requirements on companies' use of algorithms for consequential decisions.

New York City regulates the use of automated employment decision tools, requiring independent audits for bias and providing notice to job candidates.

A number of states have introduced legislation in this space, including New York, Connecticut, and Texas. The first comprehensive state-level regulation has come in Colorado. Signed into law May 17, 2024, that law places requirements on developers and deployers to use reasonable care to protect consumers from the risks of algorithmic discrimination. The bill has numerous commonalities with this bill and its predecessors.

The Virginia Legislature recently passed the High-Risk Artificial Intelligence Developer and Deployer Act, a comprehensive artificial intelligence bill focused on preventing algorithmic discrimination. However, on March 24, 2025, Virginia Governor Glenn Youngkin vetoed it.

Here in California, SB 420 (Padilla, 2025) would regulate the use of "high-risk" ADS. The bill contains several exceptions to its definition of ADS, exempts entities with 50 or fewer employees, allows state deployers to opt out of impact assessments, allows ADS to be used for two years without an impact assessment, does not enable subjects to opt out of the use of ADS, does not require audits, and provides a safe harbor to cure violations. It is supported by the Church State Council. A coalition of industry associations opposes.

SB 7 (McNerney, 2025) also regulates the use of ADS in the employment context, providing a minimum of 30 days' notice to workers before deployment as well as post-deployment rights to notice, to correct information, and to appeal. It also restricts certain uses of ADS in the employment context. SB 7 appears broadly compatible with this bill.

At the administrative level, the California Privacy Protection Agency is engaged in an ongoing rulemaking on automated decisionmaking technology. This bill provides that deployers that are businesses subject to the Privacy's agency's jurisdiction may be subject to duly-adopted privacy-related regulations.

The Civil Rights Department recently finalized a rulemaking that clarifies how California antidiscrimination protections in the employment context apply to automated decision systems.

10) **Opposition concerns.** A coalition of industry trade organizations, led by California Chamber of Commerce, details its primary concerns:

> **The sweeping scope of businesses and industries captured, the inclusion of low-risk ADS applications, and establishment of new standards/grounds for discrimination, all have to be addressed to avoid creating a de-facto restraint on technology under the guise of an impact assessment bill**. To start, businesses of 100 employees or less must be exempted. Fundamentally, both the scope and various key terms (in particular, "consequential decisions", and "automated decision systems", and "covered automated decision systems", but also other terms such as "employment-related decisions"), all require

additional clarity and/or narrowing. And insofar as the bill alters what constitutes discrimination, instead of ensuring that developers and deployers conduct evaluation and assessments that will promote responsible use of technologies to avoid violation of existing discrimination laws, the proposed change to the Unruh Civil Rights Act and introduction of new codified standards around "disparate impact" and "disparate treatment" should be fully stricken from the bill.

As to the size of the entities subject to the bill, small start-ups with a handful of employees may create widely-used ADS. Instead of focusing on the size of an entity, this bill's requirement that deployers contract with auditors to perform impact assessments is limited to deployers whose ADS impacts more than 6,000 subjects over the course of a three-year period.

As to the definitions, these are largely in harmony with related legislative efforts described above. But the terms are broad and arguably encompass technologies that pose little threat of discrimination or that play a relatively minor role in the decision-making process. While it can be argued that it is better to err on the side of inclusion, it is essential to carefully calibrate the scope of the bill. The issue warrants continued discussion with stakeholders.

Finally, as to the Unruh Act, the bill provides that a developer or deployer's compliance with the bill is relevant to, but not conclusive of, such actions. This is an evidentiary truism meant to clarify that the bill does not alter how existing anti-discrimination laws work. Rather, the bill complements these laws with a rigorous testing process that provides incentives for harms to be proactively mitigated and creates a documentary record that can be used in potential lawsuits.

The coalition continues:

Untenable opt-out and pre-and post-decision notice obligations must be deleted. All opt out and notice obligations, including the right to appeal which will have drastically different impact depending on context, must be deleted in full to make the bill workable. Such requirements are not only largely unworkable in many contexts, but also wholly unrelated to and unnecessary for there to be a bill that would require evaluations/assessments that to help reduce bias and discriminatory outcomes from the development and deployment of such ADS.

The bill recognizes that notice and opt-out provisions may be unworkable in certain situations, such as a medical emergency. Additional exceptions may need to be added. But it is not clear why these are categorically unworkable and must be deleted in full. A right to appeal, on the other hand, raises more substantial workability questions. In many cases, it appears that it is not feasible to rectify a decision, such as when a job was given to another person as a result of a discriminatory ADS. That said, the bill allows for the deployer to deny a request for appeal as long as they provide an explanation of the basis for the denial to the subject appealing the decision.

The coalition continues:

Third-party auditing requirements should be deleted in favor of self-assessments, particularly when third-party auditing would effectively grant a monopoly today and creates a costly cottage industry tomorrow, with access to highly sensitive and proprietary information, and no standards of care or liability. The third-party auditor

> requirement imposes excessive and unnecessary costs on businesses, increasing IP risks and operational inefficiencies, without providing any added consumer protections. In fact, the requirement will have the opposite effect by driving up costs for consumers. With a limited number of auditors available, a legal mandate would create a surge in demand, allowing existing auditors to charge inflated fees without competition. As businesses absorb these new compliance costs, they will be forced to raise prices, ultimately burdening consumers, reducing sales, and hindering economic growth in California. Furthermore, this requirement exceeds the scope of other U.S. laws and proposals, making California the most expensive jurisdiction for compliance. Given such impacts, the third-party auditor requirement for both developers and deployers must be removed to make this bill remotely viable.

As described above, AI governance programs such as the *Blueprint* generally call for independent third-party evaluation rather than simply entrusting entities to self-regulate. Opponents are right that an audit can be costly; hence audits are not required for deployers who use systems that impact less than 6,000 subjects over the course of a three year period. The timing of audits and scope of entities subject to those audits is an issue that warrants continued discussion with stakeholders.

> **A single enforcer is necessary to promote consistency and economic stability. Statutory penalties are far too heavy handed for a violation that involves no actual harm.** Allowing multiple enforcement entities will invariably create confusion for compliance, whereas the provision of a single enforcement entity (the Attorney General) will promote consistent interpretation and application across the state. Whereas the penalties in AB 2930 were $25,000 per violation for algorithmic discrimination only, the current penalty structure in the bill could feasibly have devastating impacts on businesses, big and most certainly small. Specifically, it allows for courts to grant plaintiffs a civil penalty of up to $25,000 per violation, without adequate clarity as to what constitutes each violation.

This bill enables civil enforcement by the Attorney General, local public prosecutors, the Civil Rights Department, and the Labor Commissioner with respect to employment-related issues. Notably, the bill, unlike AB 2930, does not provide for administrative enforcement. Further clarifications as to what constitutes a violation may be in order.

Finally, the coalition adds:

> **Preemption protections are needed, and current regulatory activities must be addressed**. Preemption is needed both to prevent conflict with localities and, similarly, to prevent concerns around state departments and agencies over-regulating this technology and getting ahead of the Legislature and Governor. These issues are too important to Californians across the state and our struggling economy to significantly delegate to unelected officials.

As local governments and administrative agencies consider their own approach to this issue, there is the looming possibility of multiple overlapping sets of ADS regulations. There is virtue in uniformity, provided that the uniform standard is sufficiently strong. The author may wish to continue exploring this issue.

ARGUMENTS IN SUPPORT: A broad coalition of civil society and labor organizations jointly write:

While the advent of generative AI and large language models has been a new piece of the puzzle, ADS have long existed in our communities. ADS have been woven into the daily lives of our community members—increasingly these tools are dictating access to and the quality of housing, healthcare, employment, credit, and many other critical services Californians need. The potential harms and biases of these systems have been well documented—from banks using lending models that were twice as likely to deny Black applicants compared to White applicants with the same financial profile, to healthcare providers using ADS that significantly underestimated the healthcare needs of Black patients compared to White ones. Technology should not be a pass to violate our civil and labor rights, this legislation regulates the misuse of automated systems.

AB 1018 would enact common-sense guardrails to help ensure that developers and deployers of these tools are testing for discriminatory outcomes prior to utilizing the tool and ensuring that consumers have the information they need to understand the role that an ADS is playing in critical decisions and what rights they have when these systems impact critical areas of their lives. Specifically, the legislation:

1. **Requires these tools to be tested before they are used on the public:** Requires that people who make and use these tools test them to make sure they do not create harm and comply with our existing rights to non-discrimination before they are sold and used on the public. It ensures that these tests are verified by an independent third party.

2. **Provides a notice to people that this tool will be used to make a critical decision about their life:** Provides people the information they need to understand where these tools are showing up in their lives and how they'll be used to determine their housing, healthcare, and job outcomes.

3. **Provides an explanation to people who were subject to a decision made with these tools:** Provides people an explanation of what the tool did, what personal information it used about them to make the decision and what role the tool played in making the decision.

4. **Ensures that every day people have more control over how these tools are used in their daily lives:** Through this bill, people will have the right to opt out of the use of an ADS tool in a critical decision about them; they will be able to correct information that the tool used to make the decision if it is inaccurate; and they will have the right to appeal the decision.

Pre-deployment testing protects Californians by ensuring that businesses understand the risks and accuracy of the systems they're purchasing and deploying to make consequential decisions about Californians. The strong notice and explanation provisions in this bill are a critical first step for consumers to know how these tools are being used and to be able to exercise their right to correct their data or appeal a decision. These protections are more important than ever as we see a Federal pullback from safety guardrails on AI and as other states and jurisdictions pass laws recognizing the danger inherent in letting these systems enter into our lives without clear protections and safeguards for the public.

ARGUMENTS IN OPPOSITION: A coalition of industry trade organizations, led by California Chamber of Commerce, writes that the bill:

. . .fails to focus on the stated objective of requirements on high-risk automated decision systems (ADS) to the detriment of every industry in the state of California, and from small business to large. Instead, the bill broadly targets businesses of all sizes, across every industry, and regulates even lower risk applications of ADS, including those that are already in use, and unnecessarily alters what constitutes discrimination under California law. Such overreach not only exposes smaller businesses to significant – if not devastating – liability even for mere errors that caused no harm to consumers, but it would also hinder many beneficial uses of ADS, including but not limited to: enabling faster approvals and expanded access to credit; enhancing real-time fraud detection; fostering job creation and new industries; improving efficiency to help level the playing field between small and large businesses; addressing major societal challenges such as bias and discrimination, economic inequality, climate change, injustices in the criminal system, disaster relief, and humanitarian aid; and advancing new treatments for previously incurable diseases.

To be clear, whether decisions are being made by humans from start to end, or a byproduct of using or incorporating new technologies in the decision-making process, we take our responsibility to not discriminate with the utmost seriousness. We believe that algorithmic discrimination, or discrimination that results from AI-enabled technologies, is already prohibited under our anti-discrimination laws because our laws are rights-based and not technology-specific. We also agree that companies need to take care to reduce bias and discrimination in decisions that have legal impact on the provision or denial of fundamental rights or essential opportunities and services, which is why so many of them already conduct impact assessments.

But it would be incredibly short-sighted for regulation to stifle innovation when alternative (human-decision driven) systems may be equally, if not more, flawed, and when properly developed and deployed ADS can enhance fairness and accountability. While ADS may pose unique challenges in terms of bias, they also pose unique advantages to combatting it as well: for example, ADS decision-making processes can be more transparent and traceable than the exercise of human discretion. The structured nature of these tools offer opportunities for detection and correction that can be more challenging to come by in human decision-making processes where human bias is often more subtle and harder to detect. Unfortunately, we believe **AB 1018** will have an undesired chilling effect on the technology and make it that much harder to develop the very tools that can help combat bias in decisionmaking.

As introduced, **AB 1018** goes far beyond ensuring developers and deployers of these technologies act responsibly to adhere to existing anti-discrimination protections and veers into an indirect restriction upon the usage of the technology itself, discouraging technological innovation and usage by making it so onerous and risky that businesses are realistically pushed back toward the alternative, human driven process – which we know based on historical evidence will not help eliminate bias and discrimination in any way.

REGISTERED SUPPORT / OPPOSITION:

Support

California State Council of Service Employees International Union (seiu California) (Co-Sponsor)
Techequity Action (Co-Sponsor)

Acce Action (alliance of Californians for Community Empowerment)
American Federation of Musicians Local 47
American Federation of Musicians, Local 7
Asian Americans Advancing Justice Southern California
California Center for Movement Legal Services
California Employment Lawyers Association
California Federation of Labor Unions, Afl-cio
California Immigrant Policy Center
California Initiative on Technology and Democracy
California National Organization for Women
California Professional Firefighters
California School Employees Association
California Women's Law Center
CFT – a Union of Educators & Classified Professionals, Aft, Afl-cio
Citizen's Privacy Coalition
Common Sense Media
Consumer Attorneys of California
Consumer Federation of America
Consumer Federation of California
Consumer Reports
Courage California
East Bay Community Law Center
Economic Security California Action
Electronic Privacy Information Center (EPIC)
End Child Poverty California Powered by Grace
Equal Rights Advocates
Fair Housing Advocates of Northern California
Felony Murder Elimination Project
Greenlining Institute
Justice2jobs Coalition
Kapor Center
LA Defensa
Laane (los Angeles Alliance for a New Economy)
Lawyers Committee for Civil Rights of the San Francisco Bay Area
Northern California District Council of the International Longshore and Warehouse Union (ILWU)
Oakland Privacy
Orange County Employees Association
Parent Voices
Powerswitch Action
Privacy Rights Clearinghouse
Public Citizen
Restoring Hope California
Rubicon Programs
Smart Justice California
Surveillance Resistance Lab
Transparency Coalition.ai
Udw/afscme Local 3930
Ufcw - Western States Council

Vision Y Compromiso
Women's Foundation California
Working Partnerships USA
Worksafe

Opposition

American Council of Life Insurers
American Property Casualty Insurance Association
American Staffing Association
Association of California Life & Health Insurance Companies
Association of National Advertisers
California Chamber of Commerce
Chamber of Progress
Civil Justice Association of California (CJAC)
College Board
Computer and Communications Industry Association
Consumer Data Industry Association
Independent Insurance Agents & Brokers of California, INC.
Internet.works
National Association of Mutual Insurance Companies
Pacific Association of Domestic Insurance Companies
Personal Insurance Federation of California
Public Risk Innovation, Solutions, and Management (PRISM)
Security Industry Association
Silicon Valley Leadership Group
Software Information Industry Association
Techca
Technet

Oppose Unless Amended

Advanced Medical Technology Association (ADVAMED)
America's Physician Groups
American College of Obstetricians & Gynecologists - District Ix
Calbroadband
Calcom Association
California Association of Health Plans
California Bankers Association
California Community Banking Network
California Credit Union League
California Hospital Association
California Life Sciences
California Medical Association (CMA)
California Mortgage Bankers Association
California Radiological Society
CTIA
Kaiser Permanente
Mortgage Bankers Association

Securities Industry and Financial Markets Association
Sutter Health

Analysis Prepared by: Josh Tosney / P. & C.P. / (916) 319-2200