SENATE JUDICIARY COMMITTEE
Senator Thomas Umberg, Chair
2023-2024 Regular Session

AB 2930 (Bauer-Kahan)
Version: June 24, 2024
Hearing Date: July 2, 2024
Fiscal: Yes
Urgency: No
CK

SUBJECT

Automated decision tools

DIGEST

This bill regulates the use of "automated decision tools" (ADTs) in order to prevent "algorithmic discrimination." This includes requirements on developers and deployers that make and use these tools to make consequential decisions to perform impact assessments on ADTs. The bill establishes the right of individuals to know when an ADT is being used, the right to opt out of its use, and an explanation of how it is used.

EXECUTIVE SUMMARY

Owing to recent advances in processing power and the rise of big data, artificial intelligence's (AI) capacity and the scope of its applications have expanded rapidly, impacting how we communicate, interact, entertain ourselves, travel, transact business, and consume media. It has been used to accelerate productivity and achieve efficiencies, but has also been used to constrain personal autonomy, compromise privacy and security, foment social upheaval, exacerbate inequality, spread misinformation, and subvert democracy.

Automated decisionmaking is one particular area where this technology is being increasingly deployed. Major transparency and fairness concerns have been raised about the use of ADTs to make consequential decisions, essentially determinations with significant legal or other material effect on one's life. This bill seeks to regulate their use by both public and private actors by requiring impact assessments to evaluate their purpose, use of data, potential for bias, and the steps taken to address those risks. The bill also ensures that individuals that are subject to ADTs know when the tool is being used to make a "consequential decision" about them, are able to opt out of their use, and are given a reasonable explanation for the automated decision made and a chance to correct any incorrect data.

The bill is author-sponsored. It is supported by various organizations, including TechEquity Action and Legal Aid at Work. It is opposed by various industry associations, including Google and the American Council of Life Insurers.

PROPOSED CHANGES TO THE LAW

Existing law:

1) Establishes the CCPA, which grants consumers certain rights with regard to their personal information, including enhanced notice, access, and disclosure; the right to deletion; the right to restrict the sale of information; and protection from discrimination for exercising these rights. It places attendant obligations on businesses to respect those rights. (Civ. Code § 1798.100 et seq.)

2) Establishes the Consumer Privacy Rights Act (CPRA), which amends the CCPA and creates the California Privacy Protection Agency (PPA), which is charged with implementing these privacy laws, promulgating regulations, and carrying out enforcement actions. (Civ. Code § 1798.100 et seq.; Proposition 24 (2020).)

3) Requires the Attorney General to adopt regulations governing access and opt-out rights with respect to businesses' use of automated decisionmaking technology, including profiling and requiring businesses' response to access requests to include meaningful information about the logic involved in those decisionmaking processes, as well as a description of the likely outcome of the process with respect to the consumer. (Civ. Code § 1798.185(a)(16).)

4) Provides that beginning the later of July 1, 2021, or six months after the PPA provides notice to the Attorney General that it is prepared to begin rulemaking, the authority assigned to the Attorney General to adopt regulations under this section shall be exercised by the PPA. (Civ. Code § 1798.185(d).)

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

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

7) Establishes the Unruh Civil Rights Act. (Civ. Code § 51.)

8) 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).)

This bill:

1) Requires a deployer to perform an impact assessment on any ADT before the tool is first deployed and annually thereafter. With respect to an ADT that a deployer first used prior to January 1, 2025, the deployer shall perform an impact assessment on that ADT before January 1, 2026, and annually thereafter.

2) Provides, notwithstanding the above, that a deployer is not required to perform an impact assessment on an ADT before using it if all of the following are true:
   a) The deployer uses the ADT only for its intended use as determined by the developer of the ADT.
   b) The deployer does not make any substantial modifications to the ADT.
   c) The developer has performed any required impact assessment on the ADT.
   d) The developer of the ADT has provided documentation to the deployer, as specified.

3) Requires a deployer to ensure that the above impact assessment includes all of the following:
   a) A statement of the purpose of the ADT and its intended benefits, uses, and deployment contexts.
   b) A description of specified features of the ADT, including the personal characteristics or attributes that the ADT will measure or assess, the method for doing so, and how they are relevant to the consequential decisions for which the ADT will be used, as well as information on its outputs.
   c) A summary of the categories of information collected from natural persons and processed by the ADT when it is used to make, or be a substantial factor in making, a consequential decision, including categories of sensitive information and information related to a natural person's receipt of sensitive services.
   d) A statement of the extent to which the deployer's use of the ADT is consistent with or varies from the statement required of the developer.
   e) An analysis of the risk of algorithmic discrimination, including adverse impacts on the basis of specified protected categories, resulting from the deployer's use of the ADT.
   f) A description of the safeguards implemented, or that will be implemented, to address reasonably foreseeable risks of algorithmic discrimination that address all of the following:
      i. Whether the ADT could be modified to mitigate the risk of algorithmic discrimination.

ii. Whether effective accommodations can be provided for any limitations on accessibility.
iii. Whether less discriminatory procedures or methods could be employed to mitigate the risk of algorithmic discrimination.

g) A description of how the ADT will be used by a natural person, or be monitored when it is used autonomously, to make, or be a substantial factor in making, a consequential decision.
h) A description of how the ADT has been or will be evaluated for validity, reliability, and relevance.

4) Requires a developer, before making an ADT that it designs, codes, or produces available to potential deployers, to perform an impact assessment on the ADT and annually thereafter. For an ADT first made available before January 1, 2025, the developer shall perform an impact assessment before January 1, 2026, and annually thereafter. The impact assessment must include all of the following:
   a) A statement of the purpose of the ADT and its intended benefits, uses, and deployment contexts.
   b) A description of the ADT's outputs and how they are used to make, or be a substantial factor in making, a consequential decision.
   c) A summary of the categories of information collected from natural persons and processed by the ADT in connection with consequential decisionmaking.
   d) An analysis of the risk of algorithmic discrimination, including adverse impacts on the basis of specific protected categories resulting from the deployer's use of the ADT.
   e) A description of the measures taken by the developer to mitigate the risk of algorithmic discrimination.
   f) A description of how the ADT can be used by a natural person, or be monitored when it is used autonomously, to make, or be a substantial factor in making, a consequential decision.
   g) A description of how the ADT has been evaluated for validity, reliability, and relevance.

5) Requires a deployer or developer to perform, as soon as feasible, an impact assessment with respect to a substantial modification to an ADT.

6) Exempts deployers with fewer than 55 employees unless they used ADTs impacting more than 999 people during the previous calendar year.

7) Requires a deployer, prior to an ADT making a consequential decision, or being a substantial factor in making a consequential decision, to notify any natural person that is subject to the consequential decision that an ADT is being used and to provide all of the following:
   a) A statement of the purpose of the ADT.

b) Contact information for the deployer.
c) A plain language description of the ADT that includes specified information, including the personal characteristics or attributes that the ADT will measure or assess, the methods by which it does so, and how those contribute to the ultimate consequential decision. The deployer must all disclose a summary of the most recent impact assessment and information on the ADT's outputs, their format, structure, and how they are used.
d) Information sufficient to enable the natural person to request to be subject to an alternative selection process or accommodation, as applicable, in lieu of the ADT, as provided.

8) Requires a deployer, if a consequential decision is made solely based on the output of an ADT, to, if technically feasible, accommodate a natural person's request to not be subject to the ADT and to instead be subject to an alternative selection process or accommodation.

9) Provides that after such a request a deployer may reasonably request, collect, and process information from a natural person for the purposes of identifying the person and the associated consequential decision. If the person does not provide that information, the deployer is not obligated to provide the alternative.

10) Requires a deployer that has deployed an ADT, to make, or be a substantial factor in making, a consequential decision concerning a natural person, to provide the person all of the following:
    a) A simple and actionable explanation that identifies the principal factors, characteristics, logic and other information related to the individual that led to the consequential decision.
    b) The role that the ADT played in the decisionmaking process.
    c) The opportunity to correct any incorrect personal data that the ADT processed in making, or as a substantial factor in making, the consequential decision.

11) Requires the notices and other communications described above to meet specified conditions, including that they be in clear and plain language in specified languages.

12) Requires a developer to provide a deployer with the results of any impact assessment performed on an ADT that the developer sells, licenses, or otherwise transfers to the deployer, along with documentation describing all the following:
    a) The intended uses and known limitations of the ADT, including any reasonably foreseeable risks of algorithmic discrimination arising from its intended use.
    b) The type of data used to program or train the ADT.

c) How the ADT was evaluated for validity and explainability.
d) The deployer's responsibilities herein and any technical information necessary for a deployer to fulfill their obligations.

13) States that the above does not require the disclosure of trade secrets, as defined in Section 3426.1 of the Civil Code.

14) Requires a deployer or developer to establish, document, implement, and maintain a governance program that contains reasonable administrative and technical safeguards designed to map, measure, and manage the reasonably foreseeable risks of algorithmic discrimination associated with the use or intended use of an ADT, as specified. This program must be designed, as provided, including the designation of at least one employee to be responsible for overseeing and maintaining the governance program and overall compliance that has the authority to assert to the employee's employer a good faith belief that the design, production, or use of an ADT fails to comply hereby and to complete assessments of any compliance issue raised by that employee. The program shall provide for annual and comprehensive reviews of policies, practices, and procedures to ensure compliance.

15) Exempts from the preceding obligation deployers with fewer than 55 employees, unless the deployer used an ADT that impacted more than 999 people during the previous calendar year.

16) Requires a deployer and developer to make publicly available, in a readily accessible manner, a clear policy that provides a summary of the types of ADTs currently in use or made available to others and how they manage the reasonably foreseeable risks of algorithmic discrimination that may arise from the use of the ADTs it currently uses or makes available to others.

17) Provides that if an impact assessment performed by a deployer identifies a reasonable risk of algorithmic discrimination, the deployer shall not use the ADT until the risk has been mitigated. If an impact assessment performed by a developer identifies such a risk under deployment conditions reasonably likely to occur in this state, the developer shall not make the ADT available to potential deployers until the risk has been mitigated.

18) Requires a state government deployer to provide, by January 1, 2026, the PPA a list of ADTs initially deployed prior to January 1, 2025, identifying:
a) Each ADT deployed and the role of each in making consequential decisions.
b) The population affected by each ADT.

19) Requires the PPA, by January 1, 2027, to establish a staggered schedule that identifies when each state government deployer shall comply, prioritizing ADTs with the highest risk for algorithmic discrimination, including civil rights violations and other discriminatory outcomes. The schedule shall require full compliance by each state deployer by January 1, 2031 with deployers in violation subject to enforcement actions by the PPA.

20) Authorizes the Civil Rights Department (CRD) to investigate a report of algorithmic discrimination or any other violation hereof.

21) Requires a deployer or a developer, upon receiving a request from the PPA, to, within 30 days of the request, provide the PPA any impact assessment that it performed pursuant hereto. The disclosure does not constitute a waiver of any attorney-client privilege or work-product protection that might otherwise exist, and the assessment is exempt from the California Public Records Act.

22) Provides that no provision herein shall be construed to require the disclosure of trade secrets, as defined.

23) Subjects a deployer or developer who violates this subdivision to liability for an administrative fine of not more than $10,000 per violation in an administrative enforcement action brought by the PPA. Each day on which an ADT is used for which an impact assessment has not been submitted shall give rise to a distinct violation.

24) Authorizes the PPA to provide an impact assessment it receives to specified public prosecutors or CRD to assist that entity in initiating or litigating a civil action.

25) Provides that, in an action brought by those entities, a court may award injunctive and declaratory relief, as well as attorneys' fees and costs. In an action for a violation involving algorithmic discrimination, a civil penalty of $25,000 per violation may also be awarded.

26) Provides a 45-day right to cure to developers and deployers.

27) Makes it unlawful for a deployer, state government deployer, or developer to retaliate against a natural person for that person's exercise of rights provided herein.

28) Exempts cybersecurity-related technology, including technology designed to detect, protect against, or respond to security incidents, identity theft, fraud, harassment, malicious or deceptive activities or any illegal activity, preserve the

integrity or security of systems, or investigate, report, or prosecute those responsible for those actions, from its provisions.

29) Clarifies that the rights, remedies, and penalties established herein are cumulative and shall not be construed to supersede the rights, remedies, or penalties established under other laws.

30) Defines the relevant terms, including:

a) "Algorithmic discrimination" means the condition in which an ADT contributes to unlawful discrimination, including differential treatment or impacts disfavoring people based on their actual or perceived race, color, ethnicity, sex, religion, age, national origin, limited English proficiency, disability, veteran status, genetic information, reproductive health, or any other classification protected by state or federal law.
b) "Automated decision tool" means an artificial intelligence system or service that makes a consequential decision, or is a substantial factor in making consequential decisions.
c) "Consequential decision" means a decision or judgment that has a legal, material, or similarly significant effect on an individual's life relating to access to government benefits or services, assignments of penalties by government, or the impact of, access to, or the cost, terms, or availability of, specified goods, services, and opportunities, including housing, employment, education, financial services, and specified aspects of the criminal justice system.
d) "Deployer" means a person, partnership, local government agency, developer, corporation, or any contractor or agent of those entities, that uses an ADT to make a consequential decision.
e) "Developer" means a person, partnership, state or local government agency, or corporation that designs, codes, or produces an ADT, or substantially modifies an artificial intelligence system or service for the intended purpose of making, or being a substantial factor in making, consequential decisions, whether for its own use or for use by a third party.
f) "Substantial factor" means an element of a decisionmaking process that is capable of altering the outcome of the process.
g) "Substantial modification" means a new version, new release, or other update to an ADT that materially changes its uses, intended uses, or outcomes.
h) "Unlawful discrimination" means any act that violates Section 51 of the Civil Code, any act that constitutes an unlawful practice or unlawful employment practice, as specified, or any other practice or act that otherwise violates a state or federal law against discrimination.

COMMENTS

1. Frameworks for responsible development and accountability in AI

As directed by the National AI Initiative Act of 2020, the National Institute of Standards and Technology (NIST) developed the AI Risk Management Framework to assist entities designing, developing, deploying, and using AI systems to help manage the many risks of AI and promote trustworthy and responsible development and use of AI systems. That framework highlights the serious risks at play and the uniquely challenging nature of addressing them in this context:

> Artificial intelligence (AI) technologies have significant potential to transform society and people's lives – from commerce and health to transportation and cybersecurity to the environment and our planet. AI technologies can drive inclusive economic growth and support scientific advancements that improve the conditions of our world. AI technologies, however, also pose risks that can negatively impact individuals, groups, organizations, communities, society, the environment, and the planet. Like risks for other types of technology, AI risks can emerge in a variety of ways and can be characterized as long- or short-term, high or low-probability, systemic or localized, and high- or low-impact.
>
> While there are myriad standards and best practices to help organizations mitigate the risks of traditional software or information-based systems, the risks posed by AI systems are in many ways unique. AI systems, for example, may be trained on data that can change over time, sometimes significantly and unexpectedly, affecting system functionality and trustworthiness in ways that are hard to understand. AI systems and the contexts in which they are deployed are frequently complex, making it difficult to detect and respond to failures when they occur. AI systems are inherently socio-technical in nature, meaning they are influenced by societal dynamics and human behavior. AI risks – and benefits – can emerge from the interplay of technical aspects combined with societal factors related to how a system is used, its interactions with other AI systems, who operates it, and the social context in which it is deployed.
>
> These risks make AI a uniquely challenging technology to deploy and utilize both for organizations and within society. [. . .]
>
> AI risk management is a key component of responsible development and use of AI systems. Responsible AI practices can help align the decisions about AI system design, development, and uses with intended aim and values. Core concepts in responsible AI emphasize human centricity, social responsibility, and sustainability. AI risk management can drive

> responsible uses and practices by prompting organizations and their internal teams who design, develop, and deploy AI to think more critically about context and potential or unexpected negative and positive impacts. Understanding and managing the risks of AI systems will help to enhance trustworthiness, and in turn, cultivate public trust.

With recent dramatic advances in the capabilities of AI systems, the need for such frameworks for accountability and responsible development have become ever more urgent. This is especially true with respect to ADTs. A number of examples of discriminatory ADTs have been provided by the author.

For instance, Amazon deployed ADTs for hiring purposes and provided an example of how bias can be built into ADTs:

> Amazon.com Inc's machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.
>
> The team had been building computer programs since 2014 to review job applicants' resumes with the aim of mechanizing the search for top talent, five people familiar with the effort told Reuters.
>
> Automation has been key to Amazon's e-commerce dominance, be it inside warehouses or driving pricing decisions. The company's experimental hiring tool used artificial intelligence to give job candidates scores ranging from one to five stars - much like shoppers rate products on Amazon, some of the people said.
>
> "Everyone wanted this holy grail," one of the people said. "They literally wanted it to be an engine where I'm going to give you 100 resumes, it will spit out the top five, and we'll hire those."
>
> But by 2015, the company realized its new system was not rating candidates for software developer jobs and other technical posts in a gender-neutral way.
>
> That is because Amazon's computer models were 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 male dominance across the tech industry.[1]

[1] Jeffrey Dastin, *Insight - Amazon scraps secret AI recruiting tool that showed bias against women* (October 10, 2018) Reuters, https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G/. All internet citations are current as of June 30, 2024.

An additional example involved a discrimination charge by the Department of Housing and Urban Development against Meta, which emphasized the importance of knowing which characteristics are being considered by an ADT.[2] The settled claims involved Meta targeting users with "housing ads based on algorithms that relied partly on characteristics protected under the Fair Housing Act, like race, national origin and sex." The charged also alleged that "Meta's lookalike or special ad audience tool allowed advertisers to target users based on protected traits."

In response to growing concerns about the increased deployment of ever-advanced ADTs, the Biden Administration published its *Blueprint for an AI Bill of Rights*, which is a set of five principles and associated practices to help guide the design, use, and deployment of AI to protect the rights of the American public:

- *Safe and Effective Systems*: You should be protected from unsafe or ineffective systems. Automated systems should be developed with consultation from diverse communities, stakeholders, and domain experts to identify concerns, risks, and potential impacts of the system.

- *Algorithmic Discrimination Protections*: Designers, developers, and deployers of automated systems should take proactive and continuous measures to protect individuals and communities from algorithmic discrimination and to use and design systems in an equitable way. This protection should include proactive equity assessments as part of the system design, use of representative data and protection against proxies for demographic features, ensuring accessibility for people with disabilities in design and development, pre-deployment and ongoing disparity testing and mitigation, and clear organizational oversight.

- *Data Privacy*: You should be protected from abusive data practices via built-in protections and you should have agency over how data about you is used. You should be protected from violations of privacy through design choices that ensure such protections are included by default, including ensuring that data collection conforms to reasonable expectations and that only data strictly necessary for the specific context is collected. Designers, developers, and deployers of automated systems should seek your permission and respect your decisions regarding collection, use, access, transfer, and deletion of your data in appropriate ways and to the greatest extent possible; where not possible, alternative privacy by design safeguards should be used. Systems should not employ user experience and design decisions that obfuscate user choice or burden users with defaults that are privacy invasive. Consent should only be used to justify collection of data in cases where it can be appropriately and

[2] Lauren Feiner, *DOJ settles lawsuit with Facebook over allegedly discriminatory housing advertising* (June 21, 2022) CNBC, https://www.cnbc.com/2022/06/21/doj-settles-with-facebook-over-allegedly-discriminatory-housing-ads.html.

meaningfully given. Any consent requests should be brief, be understandable in plain language, and give you agency over data collection and the specific context of use; current hard-to-understand notice-and-choice practices for broad uses of data should be changed. Enhanced protections and restrictions for data and inferences related to sensitive domains, including health, work, education, criminal justice, and finance, and for data pertaining to youth should put you first. In sensitive domains, your data and related inferences should only be used for necessary functions, and you should be protected by ethical review and use prohibitions. You and your communities should be free from unchecked surveillance; surveillance technologies should be subject to heightened oversight that includes at least pre-deployment assessment of their potential harms and scope limits to protect privacy and civil liberties. Continuous surveillance and monitoring should not be used in education, work, housing, or in other contexts where the use of such surveillance technologies is likely to limit rights, opportunities, or access. Whenever possible, you should have access to reporting that confirms your data decisions have been respected and provides an assessment of the potential impact of surveillance technologies on your rights, opportunities, or access.

- *Notice and Explanation*: You should know that an automated system is being used and understand how and why it contributes to outcomes that impact you. Designers, developers, and deployers of automated systems should provide generally accessible plain language documentation including clear descriptions of the overall system functioning and the role automation plays, notice that such systems are in use, the individual or organization responsible for the system, and explanations of outcomes that are clear, timely, and accessible. Such notice should be kept up-to-date and people impacted by the system should be notified of significant use case or key functionality changes. You should know how and why an outcome impacting you was determined by an automated system, including when the automated system is not the sole input determining the outcome.

- *Human Alternatives, Consideration, and Fallback*: You should be able to opt out from automated systems in favor of a human alternative, where appropriate. Appropriateness should be determined based on reasonable expectations in a given context and with a focus on ensuring broad accessibility and protecting the public from especially harmful impacts.[3]

[3] *Blueprint For An AI Bill Of Rights* (October 2022) Office of Science and Technology Policy, https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf.

2. Regulating ADTs based off of the Blueprint

This bill seeks to implement many of the principles laid out in the President's blueprint in an effort to holistically regulate ADTs, which are defined as AI systems or services that make consequential decisions, or are a substantial factor in making consequential decisions. "Consequential decision" means a decision or judgment that has a legal, material, or similarly significant effect on an individual's life relating to access to government benefits or services, assignments of penalties by government, or the impact of, access to, or the cost, terms, or availability of, specified goods, services, and opportunities, including housing, employment, education, financial services, and specified aspects of the criminal justice system.

According to the author:

> AB 2930 protects individuals from algorithmic discrimination by requiring developers and users to assess automated decision tools (ADTs) that make consequential decisions and mitigate any discovered biases. The use of ADT's have become very prominent within different sectors such as housing, employment, and even in criminal justice sentencing and probation decisions. The algorithms used within ADTs can be prone to unrepresentative datasets, faulty classifications, and flawed design, which can lead to biased, discriminatory, or unfair outcomes. These tools can exacerbate the harms they are intended to address and ultimately hurt the people they are supposed to help. As the use of decision making via algorithm becomes more prevalent in our daily lives, it is crucial that we take the necessary steps to ensure that they are used ethically and responsibly.

a. *Impact assessments*

In order to ensure ADTs are fair, transparent, and aligned with basic ethical standards, the bill requires developers and deployers to conduct impact assessments, laying out the details of the tool and an analysis of the risk of "algorithmic discrimination," the ultimate target of the bill. "Algorithmic discrimination" means the condition in which an ADT contributes to unlawful discrimination, including differential treatment or impacts disfavoring people based on their actual or perceived race, color, ethnicity, sex, religion, age, national origin, limited English proficiency, disability, veteran status, genetic information, reproductive health, or any other classification protected by state or federal law.

As stated in the President's Blueprint:

> *Algorithmic Discrimination Protections*: Designers, developers, and deployers of automated systems should take proactive and continuous measures to protect

> individuals and communities from algorithmic discrimination and to use and design systems in an equitable way. This protection should include proactive equity assessments as part of the system design, use of representative data and protection against proxies for demographic features, ensuring accessibility for people with disabilities in design and development, pre-deployment and ongoing disparity testing and mitigation, and clear organizational oversight.

Before making an ADT that it designs, codes, or produces available to potential deployers, a developer is required to perform such an assessment on the ADT and annually thereafter. The bill provides detailed specifications for what needs to be included in the impact assessment. This includes a statement of the purpose of the ADT and its intended benefits, uses, and deployment contexts and a description of the ADT's outputs and how they are used to make, or be a substantial factor in making, a consequential decision. The impact assessment must also provide a summary of the categories of information collected from natural persons and processed by the ADT in connection with consequential decisionmaking.

Most importantly, the impact assessment must include an analysis of the risk of algorithmic discrimination, including adverse impacts on the basis of specified protected categories resulting from the deployer's use of the ADT. A description of the measures taken by the developer to mitigate that risk and of how the ADT can be used by a natural person, or be monitored when it is used autonomously, to make, or be a substantial factor in making, a consequential decision.

The developer is then required to provide a deployer, the person or entity, including a governmental entity, that uses an ADT to make consequential decisions, with the results of any impact assessment performed on an ADT that the developer sells, licenses, or otherwise transfers to the deployer, along with documentation describing all the following:

- The intended uses and known limitations of the ADT, including any reasonably foreseeable risks of algorithmic discrimination arising from its intended use.
- The type of data used to program or train the ADT.
- How the ADT was evaluated for validity and explainability.
- The deployer's responsibilities herein and any technical information necessary for a deployer to fulfill their obligations.

Deployers must also complete similar impact assessments, unless the following conditions are met:

- The deployer uses the ADT only for its intended use as determined by the developer.
- The deployer does not make any substantial modifications to the ADT.
- The developer has performed any required impact assessment on the ADT.

- The developer of the ADT has provided the above documentation to the deployer.

If those conditions are not met, the deployer must complete an impact assessment that includes much of what the developer is required to include. The assessment must also include a description of specified features of the ADT, including the personal characteristics or attributes that the ADT will measure or assess, the method for doing so, and how they are relevant to the consequential decisions for which the ADT will be used, as well as information on its outputs. It must also include a statement of the extent to which the deployer's use of the ADT is consistent with or varies from the statement required of the developer.

These impact assessments must be performed annually and, if there is a substantial modification made to the ADT, as soon as feasible thereafter. Where the assessment identifies a reasonable risk of algorithmic discrimination, the developer shall not make it available to deployers, and the deployer shall not use the ADT, until the risk has been mitigated.

A deployer or developer must also establish, document, implement, and maintain a governance program that contains reasonable administrative and technical safeguards designed to map, measure, and manage the reasonably foreseeable risks of algorithmic discrimination associated with the use or intended use of an ADT, as specified. This requires designation of at least one employee to be responsible for overseeing and maintaining the governance program and overall compliance with the provisions of this bill. The program shall all provide for annual and comprehensive reviews of policies, practices, and procedures to ensure compliance.

To lessen the impact on small businesses, the above requirements do not apply to deployers with fewer than 55 employees unless their use of ADTs impacts 1000 people or more.

Deployers that are state government agencies must provide the PPA a list of ADTs deployed, the role they have in making consequential decisions, and the populations thereby affected. The PPA is directed to establish a schedule for agency compliance.

b. Notice, disclosures, and the right to opt out for individuals subjected to ADTs

Again, the Blueprint provides:

- *Notice and Explanation*: You should know that an automated system is being used and understand how and why it contributes to outcomes that impact you. Designers, developers, and deployers of automated systems should provide generally accessible plain language documentation including clear descriptions of the overall system functioning and the role automation plays, notice that such

systems are in use, the individual or organization responsible for the system, and explanations of outcomes that are clear, timely, and accessible. Such notice should be kept up-to-date and people impacted by the system should be notified of significant use case or key functionality changes. You should know how and why an outcome impacting you was determined by an automated system, including when the automated system is not the sole input determining the outcome.

- *Human Alternatives, Consideration, and Fallback*: You should be able to opt out from automated systems in favor of a human alternative, where appropriate. Appropriateness should be determined based on reasonable expectations in a given context and with a focus on ensuring broad accessibility and protecting the public from especially harmful impacts.

This bill effectuates these principles by requiring a deployer, prior to an ADT making a consequential decision, or being a substantial factor in making a consequential decision, to notify any natural person that is subject to the consequential decision that an ADT is being used. That person shall be provided all of the following:

- A statement of the purpose of the ADT.
- Contact information for the deployer.
- A plain language description of the ADT that includes specified information, including the personal characteristics or attributes that the ADT will measure or assess, the methods by which it does so, and how those contribute to the ultimate consequential decision. The deployer must all disclose a summary of the most recent impact assessment and information on the ADT's outputs, their format, structure, and how they are used.
- Information sufficient to enable the natural person to request to be subject to an alternative selection process or accommodation, as applicable, in lieu of the ADT, as provided.

Furthermore, the bill requires a deployer, if a consequential decision is made solely based on the output of an ADT, to accommodate a natural person's request to not be subject to the ADT and to instead be subject to an alternative selection process or accommodation However, to mitigate the impact on businesses this is only required if technically feasible and if the person does not provide specified information, the deployer is not so obligated.

Once the ADT is involved in making a consequential decision, the deployer is required to provide the person all of the following:

- A simple and actionable explanation that identifies the principal factors, characteristics, logic and other information related to the individual that led to the consequential decision.
- The role that the ADT played in the decisionmaking process.

- The opportunity to correct any incorrect personal data that the ADT processed in making, or as a substantial factor in making, the consequential decision.

This provides transparency and some measure of control to the subject of the ADT.

The required notices and other communications must meet specified conditions, including that they be in clear and plain language in specified languages.

The bill makes it unlawful for a deployer, state government deployer, or developer to retaliate against a natural person for exercising any rights provided herein.

c. Enforcement and oversight

The PPA is the primary regulatory and enforcement entity for implementing the CCPA and CPRA. The PPA is currently in the process of drafting regulations to govern the use of automated decisionmaking technology by CCPA-covered businesses.

This bill requires deployers and developers to provide the PPA with any impact assessment performed within 30 days of a request. The PPA is authorized to bring an administrative enforcement action against a developer or deployer in violation of these provisions seeking an administrative fine of up to $10,000 per violation. CRD is explicitly granted authority to investigate a report of algorithmic discrimination or any other violation of the bill.

The bill further provides that CRD, the Attorney General, district attorneys, county counsel, city attorneys, and certain city prosecutors, as specified, may bring a civil action against a developer or deployer in violation seeking injunctive and declaratory relief, as well as attorney's fees and costs. Where the violation involves algorithmic discrimination, the public prosecutors may also seek a civil penalty of $25,000 per violation. To assist these entities in such enforcement actions, the PPA is authorized to provide them with any impact assessments it receives. However, before initiating such actions, the public prosecutor must provide the developer or deployer in alleged violation with 45-days written notice and an opportunity to cure. If the violation is cured and a written notice attesting to such cure is provided, no claim for injunctive relief may be maintained.

3. Other jurisdictions

A number of jurisdictions have stepped forward to respond to the dramatic increase in ADT usage. 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, a number of states have introduced legislation in this space, including New York and Connecticut. However, the first comprehensive state-level regulation has come in Colorado.

The Colorado law, approved by their Governor on May 17, 2024, places requirements on developers and deployers to use reasonable care to protect consumers from the risks of algorithmic discrimination. Many provisions of the Colorado law are found in this bill.

4. Stakeholder engagement and positions

A number of stakeholders from industry associations, to labor groups, to public-interest advocacy organizations have engaged with the author and this Committee on the bill. A broad set of amendments recently made to the bill respond to a number of the concerns that have been raised by various groups and represent a reasonable compromise. For instance, recent amendments replaced "unjustified differential treatment or impacts" with "unlawful discrimination" in response to concerns from businesses that the former wording lacked clarity.

One additional change was an expansion of what materials were exempt from disclosure. Formerly, only trade secrets were not subject to disclosure pursuant to the California Public Records Act. Recent amendments make all impact assessments disclosed to the PPA exempt. The bill still makes clear that nothing therein requires the disclosure of trade secrets. However, claiming exemptions for trade secrets is a heavily litigated subject with an over reliance on such clauses. That is what led Colorado to specifically provide in its law that where a developer or deployer withholds information on this basis, it must notify the relevant party, including the consumer about the withholding and the basis for it. The author has agreed to amendments that place similar protections into the bill.

Another recent amendment fleshes out an existing exemption for cybersecurity-related technology, including technology designed to detect, protect against, or respond to security incidents, identity theft, fraud, harassment, malicious or deceptive activities or any illegal activity, preserve the integrity or security of systems, or investigate, report, or prosecute those responsible for those actions. The corresponding exemption in the Colorado law, however, specifically caveats a similar exemption to the technology by extending to it unless deployed to make, or be a substantial factor in making, a consequential decision. The author may wish to consider amendments that provide a similar caveat to ensure there are no loopholes in the protections provided for by the bill.

The recent changes have brought a number of groups from a support-if-amended position into support. A coalition of groups, including Consumer Reports, Equal Rights Advocates, and the Greenlining Institute, write in support:

> While the advent of generative AI and large language models has been a new piece of the puzzle, ADTs have long existed in our communities. ADTs have been woven into the daily lives of our community members—

> increasingly these tools are dictating access to and the quality of housing, employment, credit, and many other critical services Californians need. The potential harms of these systems have also been well documented—from predictive models on recidivism that have been found to incorrectly predict that Black defendants are two times more likely to recidivate than white defendants to the use of algorithms to determine outcomes for family separations resulting in a potentially disparate impact on disabled parents and children. These tools have also been integrated into private businesses' enterprise systems and used by governments to dictate who receives unemployment insurance or to determine who may get access to affordable housing.
>
> AB 2930 would enact common-sense guardrails to help ensure that developers and deployers of these tools are obligated to test and mitigate for discriminatory outcomes prior to the sale or use of these tools in our communities. Specifically, the legislation would:
>
> 1. Require developers and deployers to conduct pre-deployment impact assessments to determine any potential for discrimination on people with protected class status;
> 2. Prohibit the sale or use of an ADT that may create a discriminatory outcome on people within a protected class until that adverse impact has been addressed and resolved;
> 3. Provide consumers with pre-use notice of the tool, a post-use explanation, the right to correct inaccurate information, and access to alternative selection procedures.

Various organizations have written requesting amendments to narrow the scope of the bill. Writing in a support if amended position, the California Nurses Association requests amendments to exclude post-hiring employment-related ADTs. Additionally, a number of associations representing various sectors have written in opposition to the bill and are requesting exemptions for their respective industries. This includes separate coalitions representing health care entities, the finance industry, and the insurance industry. The general argument is that other laws already adequately regulate these industries and the additional obligations imposed by this bill will disrupt operations. The Consumer Technology Association also writes in opposition that the focus on all ADTs is overbroad:

> AB 2930 would require impact assessments be performed for "any" ADT the deployer uses, regardless of whether the use of the tool presents any significant risks. A risk-based approach to regulating AI tools is necessary to avoid overbroad regulations and costly new mandates that can, and should, be narrowly focused on only those use cases presenting greatest risks to individuals or society.

It should be noted that the bill makes clear that it is cumulative to existing law. Therefore, the bill does not impact any other legal requirements, rights, or remedies provided under existing law. For instance, nothing therein should be interpreted to impact existing anti-discrimination laws or the enforcement of them, and compliance with the provisions of this bill are not relevant to actions brought pursuant to other laws.

Writing in opposition, Google argues:

> Section 22576(a) currently provides that "'Algorithmic discrimination' means the condition in which an automated decision tool contributes to unjustified differential treatment or impacts disfavoring people based on their actual or perceived race, color, ethnicity, sex, religion, age, national origin, limited English proficiency, disability, veteran status, genetic information, reproductive health, or any other classification protected by state law." This broad definition of algorithmic discrimination advances a novel legal concept of algorithmic discrimination that has no clear contours and is untethered from unlawful discrimination under California's existing civil rights laws.
>
> We respectfully request that at minimum AB 2930 be amended to replace the word "unjustified" with "unlawful" and "contributes to" with "results in" in the definition of algorithmic discrimination. Section 22576(a).[4]
>
> More generally, the bill should not create a two-tiered standard for illegal discrimination with one for human decisionmakers and another for software. Unless "algorithmic discrimination" is synomous [*sic*] with "unlawful discrimination", Section 22756.6's prohibitions on using and making available ADTs resulting in algorithmic discrimination would create a new category of illegal discrimination. Even if the bill were amended to add "currently unlawful" to the definition of algorithmic discrimination, the provision would create confusion by adding a new enforcement mechanism.
>
> We believe that if it is illegal to do something without AI, it is illegal to do it with AI. If there is concern that existing California civil rights statutes do not explicitly state that they apply to AI or ADTs, the better approach would be to add clarifying language to existing laws. These laws can require safe and responsible development and deployment of ADTs for purposes of preventing illegal discrimination and punishing the failure to satisfy impact assessment, governance, and transparency requirements.

[4] As referenced above, some of these concerns have been addressed by amendments recently taken.

Existing law should continue to prohibit and punish acts resulting in unlawful discrimination.

SUPPORT

American Federation of Musicians, Local 7
California Employment Lawyers Association
Center for Democracy and Technology
Center on Race and Digital Justice Secure Justice
Consumer Reports
East Bay Community Law Center
Economic Security California Action
Equal Rights Advocates
The Greenlining Institute
Legal Aid at Work
Rise Economy
Techequity Collaborative

OPPOSITION

ACLU California Action
Advanced Medical Technology Association
American Council of Life Insurers
American Property Casualty Insurance Association
America's Physician Groups
Association of California Life & Health Insurance Companies
California Association of Health Plans
California Bankers Association
California Community Banking Network
California Credit Union League
California Financial Services Association
California Hospital Association
California Life Sciences
California Medical Association
California Mortgage Bankers Association
Consumer Technology Association
Electronic Frontier Foundation
Google
Kaiser Permanente
Mortgage Bankers Association
National Association of Mutual Insurance Companies
Orange County Business Council
Pacific Association of Domestic Insurance Companies
Personal Insurance Federation of California

Sutter Health
Verizon Communications

RELATED LEGISLATION

Pending Legislation:

SB 892 (Padilla, 2024) requires the California Department of Technology (CDT) to develop and adopt regulations to create an AI risk management standard, as specified. It requires the AI risk management standard to include, among other things, a detailed risk assessment procedure for procuring automated decision systems (ADS), as defined, that analyzes specified characteristics of the ADS, methods for appropriate risk controls, as provided, and adverse incident monitoring procedures. It requires CDT to, among other things, collaborate with specified organizations to develop the AI risk management standard. SB 892 is currently in the Assembly Privacy and Consumer Protection Committee.

AB 2885 (Bauer-Kahan, 2024) establishes a uniform definition for "artificial intelligence" in California's codes. AB 2885 is currently in the Senate Appropriations Committee.

Prior Legislation:

AB 302 (Ward, Ch. 800, Stats. 2023) requires CDT, on or before September 1, 2024, to conduct a comprehensive inventory of all high-risk ADS that have been proposed for use, development, or procurement by, or are being used, developed, or procured by, any state agency.

AB 331 (Bauer-Kahan, 2023) was substantially similar to the current bill. AB 331 died in the Assembly Appropriations Committee.

PRIOR VOTES:

Assembly Floor (Ayes 50, Noes 14)
Assembly Appropriations Committee (Ayes 11, Noes 4)
Assembly Judiciary Committee (Ayes 9, Noes 2)
Assembly Privacy and Consumer Protection Committee (Ayes 8, Noes 3)
****