ACLU CALIFORNIA ACTION

P.O. Box 630
1017 L Street
Sacramento 95814

Tel: (916) 442-1036

May 6, 2024

Secretary Amy Tong
California Government Operations Agency

Director Liana Bailey-Crimmins
California Department of Technology

Director Jeffrey Marino
Office of Data and Innovation

Director Dee Dee Myers
Governor's Office of Business and Economic Development

**Sent via email**

**Re: ACLU California Action Comment on Executive Order N-12-23 regarding Generative Artificial Intelligence, Part 3(b)**

Dear Secretary Tong, Director Bailey-Crimmins, Director Mariano and Director Myers:

The American Civil Liberties California Action ("ACLU") applaud California's leadership in examining the government's deployment of artificial intelligence ("AI") technology while remaining clear-eyed about the risks such systems could pose to civil rights and vulnerable communities. The ACLU appreciates the opportunity to comment[1] on Executive Order N-12-23 on Generative Artificial Intelligence, Part 3(b) ("the Executive Order"), which directs the creation of guidelines for State agencies and departments to analyze the impact of GenAI systems on vulnerable communities, including high-risk use cases, that shall inform whether and how an agency or department deploys a particular system (hereinafter the "Guidelines").[2]

AI systems hold the promise to make life better for Californians if they are built carefully, proven effective for the task at hand, and used thoughtfully to address

---

[1] The ACLU would like to thank Bani Sapra and Nicole Bloomfield, law students in Berkeley Law's Samuelson Law, Technology & Public Policy Clinic, as well as clinic director Catherine Crump, for their assistance researching and writing this comment.

[2] Specifically, the Executive Order directs the Government Operations Agency, the California Department of Technology, and the Office of Data and Innovation to "develop guidelines for State agencies and departments to analyze the impact that adopting a GenAI tool may have on vulnerable communities, including criteria to evaluate equitable outcomes in deployment and implementation of high-risk use cases. These guidelines and criteria shall inform whether and how a State agency or department deploys a particular GenAI tool." Cal. Exec. Order N-12-23 §3(b) at 4 (Sep. 6, 2023), *available at* https://www.gov.ca.gov/wp-content/uploads/2023/09/AI-EO-No.12-_-GGN-Signed.pdf.

longstanding systemic challenges to access, equity, and justice that have disproportionately harmed marginalized Californians. But depending on the particular system and use case, AI can also magnify and expand threats to rights, health, and safety. Despite the apparent novelty of many AI systems, existing rights and principles can serve as a guide to ensure that the State's use of AI does not contribute to harm. Among the core rights implicated by AI systems is our state constitutional right to privacy in Article I, Section 1 of the California Constitution.[3] Enacted in 1972, this right guarantees an inalienable right to privacy to all Californians and protects against invasions by both government and private parties. Article I, Section 1 is intended to impose "effective restraints" on the "accelerating encroachment on personal freedom and security caused by increased surveillance and data collection activity in contemporary society."[4] This fundamental right to privacy should inform these Guidelines, including whether and how agencies and departments utilize AI systems.[5]

Crucially, the Executive Order says the Guidelines should help agencies and departments decide *whether* to deploy an AI system. This important question is the focus of the ACLU's comment. It is critical to thoughtfully consider why and whether to procure a system, before discussing how to do so. This process will help ensure that systems will actually properly address the specific or systemic issues facing Californians. Non-AI legacy systems are by no means perfect at serving Californians. This time of technological advancement is an opportunity for California to also reexamine existing systems and consider whether technology can help the government effectively address systemic problems and make life better for vulnerable Californians.

These Guidelines should ensure that whenever AI is proposed, State agencies and Departments will first identify the specific problem they want to solve in consultation with vulnerable populations and impacted people. Agencies and departments should consider both how existing non-AI solutions and AI systems can address these problems efficiently and without impacts on people's rights. They should conduct an assessment that considers the proposed systems and weighs evidence-based public benefits with public harms, particularly risks for harm to vulnerable communities. Informed by this analysis, agencies and departments should only deploy an AI system where there are demonstrable public benefits that substantially outweigh potential public harms. Ultimately, a decision about whether to deploy AI should be driven by the real needs and interests of potentially impacted communities, and not by governments and AI companies acting alone behind closed doors.

To assist in protecting vulnerable Californians, we offer the following recommendations: **First**, the Guidelines should apply not just to Generative AI ("Gen AI") but also to any automated systems that make decisions impacting the lives of

[3] Cal. Const. Art. I § 1.
[4] *Id.*; *White v. Davis*, 533 P.2d 222, 233 (1974).
[5] *Id.*

Californians. **Second**, to recognize existing rights, including the California constitutional right to privacy, and build on federal best practices, the Guidelines should adopt a decisionmaking standard that is based on community input and an AI impact assessment and requires an evidence-based demonstration that the public benefit justifies the proposed use and outweighs the potential harms. **Third**, the Guidelines should proactively recognize that some uses of AI are too harmful to be deployed by agencies under any circumstances and should prohibit agency procurement and use of technologies that pose an unacceptable risk to vulnerable communities. **Finally**, given the extensive use of AI by California local governments and agencies, we urge your office to explore assisting localities that wish to adhere to the Guidelines voluntarily, including through both technical assistance and training.

## I. The Guidelines Should Address Both Generative AI Systems and Other AI and Automated Systems.

These Guidelines present an opportunity to advance equity by helping agencies and departments decide *whether* and *how* to deploy GenAI systems by centering their potential impact on vulnerable communities. However, the proposed Guidelines are currently aimed at a narrow subset of AI systems and, as a result, fail to capture many current and foreseeable uses of automated decision systems that impact vulnerable communities.[6] Because of their ongoing and potential impact on people's lives, other automated decision-making systems should also be within the scope of the Guidelines. Building on the White House's Blueprint for an AI Bill of Rights, the Guidelines should apply to (1) "any automated system"[7] that (2) has the potential to meaningfully impact the California public's "rights, opportunities, or access to critical resources or services."[8]

The government's exploration of GenAI is necessarily a high stakes endeavor because the government holds immense power over the lives of Californians, including the power to grant or deny benefits, gatekeep access to important services, and impose financial penalties and the harsh consequences of criminal law. While California aspires to use GenAI for everything from "improv[ing] the equitable and timely delivery

[6] Cal. Exec. Order N-12-23 §3(b) at 4 (Sep. 6, 2023), *available at* https://www.gov.ca.gov/wp-content/uploads/2023/09/AI-EO-No.12-_-GGN-Signed.pdf.

[7] An "automated system" is "any system, software, or process that uses computation as whole or part of a system to determine outcomes, make or aid decisions, inform policy implementation, collect data or observations, or otherwise interact with individuals and/or communities. Automated systems include, but are not limited to, systems derived from machine learning, statistics, or other data processing or artificial intelligence techniques, and exclude passive computing infrastructure." White House Off. of Sci. & Tech. Pol'y, *Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People* at 10 (Oct. 2022), *available at* https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf.

[8] White House Off. Of Sci. & Tech. Pol'y, *Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People* at 8 (Oct. 2022), *available at* https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf.

of services"[9] to "analyz[ing] healthcare claims or tax filing data to detect fraud" to making government information and materials more accessible, the State should be clear-eyed about whether there are evidence-based public benefits for these systems and the potential harms to the public as well.[10] There are already numerous known risks that must be considered. As things stand, generative AI systems have a "tendency to create false information."[11] A separate issue, as computer scientist and AI researcher Timnit Gebru has discussed, is that large datasets used to train models "encode bias potentially damaging to marginalized populations," which "set[s] up models trained on these datasets to further amplify biases and harms."[12] As easily available human-generated training data runs out, some companies are also turning to "synthetic data," or AI-generated data, to train systems, a practice that may "only compound[] the problems of bias from the past."[13]

In the many areas where there are real-life consequences for errors and bias, California should be particularly careful when it comes to the use of GenAI. For example, in the context of processing names on benefits or claims forms, a GenAI system that "favored names from some demographics more often than others" could lead to discrimination against Californians and the deprivation of needed resources and help.[14] Even today, as California explores a pilot program to use GenAI to answer tax questions,[15] "red flags" have been raised about the accuracy of early AI offerings by

---

[9] Cal. Exec. Order N-12-23 §3(b) at 1 (Sep. 6, 2023), *available at* https://www.gov.ca.gov/wp-content/uploads/2023/09/AI-EO-No.12-_-GGN-Signed.pdf.
[10] Cal. Gov't Operations Agency, *Benefits and Risks of Generative Artificial Intelligence Report* 7-12 (Nov. 2023), *available at* https://www.govops.ca.gov/wp-content/uploads/sites/11/2023/11/GenAI-EO-1-Report_FINAL.pdf.
[11] Anika C. Navaroli, *Op-Ed: AI's Most Pressing Ethics Problem*, Columbia Journalism Review (Apr. 23, 2024), *available at* https://www.cjr.org/tow_center/op-ed-ais-most-pressing-ethics-problem.php.
[12] *Id.*; Timnit Gebru, Emily M. Bender, Angelina McMillan-Major, Shmargaret Shmitchell, *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?* 🦜, in Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency 610, 613 (2021), *available at* https://dl.acm.org/doi/10.1145/3442188.3445922.
[13] Anika C. Navaroli, *Op-Ed: AI's Most Pressing Ethics Problem*, Columbia Journalism Review (Apr. 23, 2024), *available at* https://www.cjr.org/tow_center/op-ed-ais-most-pressing-ethics-problem.php.
[14] In its test, Bloomberg asked a GenAI model to rank fictitious resumes to identify the most qualified candidate. The key difference between the largely similar resumes was the name of the fictitious candidate, and whether that name was statistically associated with men or women who are either Black, White, Hispanic, or Asian. After running the experiment 1000 times, the journalists discovered that "resumes with names distinct to Black Americans were the least likely to be ranked as the TOP CANDIDATE for a financial analyst role, compared to resumes with names associated with other races and ethnicities." Leon Yin, Davey Alba, Leonardo Nicoletti, *OpenAI's GPT Is a Recruiter's Dream Tool. Tests Show There's Racial Bias*, Bloomberg.com (Mar. 8, 2024), *available at* https://www.bloomberg.com/graphics/2024-openai-gpt-hiring-racial-discrimination/.
[15] Khari Johnson, *California Plans to Use AI to Answer Your Tax Questions*, CalMatters (Feb. 8, 2024), *available at* http://calmatters.org/economy/technology/2024/02/cdtfa-generative-ai/.

H&R Block and TurboTax that purport to do similar tasks.[16] Stanford researchers also considered the use of Gen AI in the provision of legal information and advice and concluded that "the current limitations of [large language models] pose a risk of further deepening existing legal inequalities, rather than alleviating them."[17]

The Executive Order currently focuses on GenAI systems,[18] as does the definition of GenAI in the recently-released GenAI Guidelines for Public Sector Procurement, Uses and Training.[19] A separate report mandated by the Executive Order distinguishes content-generating GenAI systems such as ChatGPT and Dall-E from "conventional AI" systems such as robotic process automation and fraud-detection tools. But both Gen AI and AI systems raise similar risks of harm to vulnerable communities, including issues of privacy, reliability, accuracy, and bias that underscore the need for thoughtful consideration prior to any acquisition and deployment decisions.

Since many forms of AI are in their infancy, including GenAI, we urge the Guidelines to be more technology neutral and address AI rather than only a narrower subset of Gen AI systems. A more inclusive definition would better serve the desired goal of ensuring that proposed AI systems are evaluated for "equitable outcomes."[20]

### A. The Guidelines Should Address Ongoing Uses of AI Systems by California Departments.

The following are two examples of California departments using AI systems. These examples illustrate why it is important for the Guidelines to address broader types of AI systems to better protect Californians, especially vulnerable communities:

---

[16] Geoffrey A. Fowler, *TurboTax and H&R Block Now Use AI for Tax Advice. It's Awful*., Washington Post (Mar. 11, 2024), *available at* https://www.washingtonpost.com/technology/2024/03/04/ai-taxes-turbotax-hrblock-chatbot/.

[17] Matthew Dahl, Varun Magesh, Mirac Suzgun, Daniel E. Ho, *Hallucinating Law: Legal Mistakes with Large Language Models are Pervasive*, Stanford University Institute for Human-Centered Artificial Intelligence (Jan. 11, 2024), *available at* https://hai.stanford.edu/news/hallucinating-law-legal-mistakes-large-language-models-are-pervasive.

[18] Cal. Exec. Order N-12-23 §3(b) at 4 (Sep. 6, 2023), *available at* https://www.gov.ca.gov/wp-content/uploads/2023/09/AI-EO-No.12-_-GGN-Signed.pdf (instructing that guidelines be drafted for State agencies and departments to "analyze the impact that adopting a GenAI tool may have on vulnerable communities.")

[19] These guidelines were drafted in response to 3(a) of the Executive Order and define GenAI as "[p]retrained AI models that can generate images, videos, audio, text, and derived synthetic content… by analyzing the structure and characteristics of the input data to generate new, synthetic content similar to the original." This definition excludes "[d]ecision support, machine learning, natural language processing/translation services, computer vision and chatbot technologies or activities support," which "may be related to GenAI, but they are not GenAI on their own." Cal. Gov't Operations Agency et al., *State of California GenAI Guidelines for Public Sector Procurement, Uses and Training* at 6 (Mar. 2024), *available at* https://cdt.ca.gov/wp-content/uploads/2024/03/3a-GenAI-Guidelines.pdf.

[20] Cal. Exec. Order N-12-23 §3(b) at 4 (Sep. 6, 2023), *available at* https://www.gov.ca.gov/wp-content/uploads/2023/09/AI-EO-No.12-_-GGN-Signed.pdf.

**First**, the California Department of Public Health's ("CDPH") used an algorithm to deliver additional vaccines to underserved communities during the Covid-19 pandemic, which had a critical impact on the health outcomes of vulnerable communities.[21] The system assigned deliveries based off ZIP codes rather than more specific census tracts, resulting in a failure to account for low-income neighborhoods based in wealthier ZIP codes.[22] The ACLU of Northern California's analysis suggested that CDPH's choice to use less precise boundaries risked overlooking more than two million vulnerable Californians.[23] As currently written, the Executive Order and its proposed Guidelines would not address the CDPH's use of automated systems such as this vaccine-allocation algorithm, despite its substantial impact on a critical government service provided to improve the health of vulnerable communities.[24]

**Second**, the Employment Development Department's ("EDD") use of AI systems to verify unemployment benefits claims also raised equity issues for vulnerable communities.[25] During the Covid-19 pandemic, the EDD used machine learning to review nearly ten million claims for "potentially fraudulent characteristics,"[26] and stopped paying unemployment benefits for 1.1 million claims that were identified as potentially fraudulent.[27] But 600,000 of the claims flagged by the fraud-detection algorithm were later confirmed as legitimate.[28] EDD also contracted with ID.me, a service that utilized face surveillance to verify unemployment benefits claims.[29] The technology often failed to identify unemployed applicants, forcing many applicants to be manually verified.[30] Since low-income workers of color were disproportionately

---

[21] *See* Jacob Snow, *California's "Equity" Algorithm Could Leave 2 Million Struggling Californians Without Additional Vaccine Supply*, ACLU of NorCal (May 6, 2021), *available at* https://www.aclunc.org/blog/californias-equity-algorithm-could-leave-2-million-struggling-californians-without-additional.
[22] *Id.*
[23] *Id.*
[24] Cal. Exec. Order N-12-23 §3(b) at 4 (Sep. 6, 2023), *available at* https://www.gov.ca.gov/wp-content/uploads/2023/09/AI-EO-No.12-  -GGN-Signed.pdf.
[25] Lauren Hepler, *Internal Documents Reveal The Story Behind California's Unemployment Crash*, CalMatters (Nov. 7, 2023), *available at* https://calmatters.org/economy/2023/11/california-unemployment-covid/.
[26] *See* Cal. Leg. Analyst's Off., *Assessing Proposals to Address Unemployment Insurance Fraud* at 4 (2022), *available at* https://lao.ca.gov/reports/2022/4542/Unemployment-Insurance-Fraud-021522.pdf
[27] *Id.*; *see also* Lauren Hepler, *Amid California's Unemployment Crisis, A Tech Gold Rush*, CalMatters (Apr. 5, 2021), *available at* https://calmatters.org/economy/2021/04/california-unemployment-crisis-contracts/.
[28] Cal. Leg. Analyst's Off., *Assessing Proposals to Address Unemployment Insurance Fraud* at 4 (2022), *available at* https://lao.ca.gov/reports/2022/4542/Unemployment-Insurance-Fraud-021522.pdf.
[29] *Id.* at 3-4.
[30] *See* Lauren Hepler, *Amid California's Unemployment Crisis, A Tech Gold Rush*, CalMatters (Apr. 5, 2021), *available at* https://calmatters.org/economy/2021/04/california-unemployment-crisis-contracts/ ("There are entire Internet subcultures dedicated to frustrated unemployment claimants… navigating hours-long waits for ID.me video calls to verify their identities.")

impacted by the job losses accompanying the pandemic,[31] the EDD's use of a fraud-detection algorithm disproportionately forced vulnerable workers to rebut false fraud allegations and mitigate wrongful denials or terminations of their benefits at an already stressful time. Furthermore, its use of face surveillance—a technology that is prone to error and often suffers from systemic racial and gender bias[32]—compounded the risk that EDD's resolution of unemployment benefits claims disproportionately impacted vulnerable communities. Despite such issues, the EDD's use of machine learning and face surveillance would not be subject to a set of Guidelines solely focused on GenAI.

### B. GenAI Systems, Other AI, and Automated Systems Raise Similar Issues.

A more inclusive definition would also allow State agencies and departments to address common risks raised by GenAI and other AI and automated systems. As the California Government Operations Agency acknowledged, GenAI both shares and amplifies the risks posed by conventional AI.[33] A variety of AI systems generally have explainability and interpretability issues, which makes it difficult to address cases where the technology produces an unexpected result that impacts the validity and consistency of the system's answers.[34] State agencies and departments using other forms of AI will thus face similar transparency and accountability challenges in attempting to disclose information about decisions that may implicate vulnerable Californians.[35] Both GenAI models and other automated systems also share the risk of perpetuating societal biases by relying on training data that reflects such biases.[36]

---

[31] Alex Bell, Thomas J. Hedin, Geoffrey Schnorr, Till von Wachter, Cal. Pol'y Lab, *An Analysis of Unemployment Insurance Claims in California During the COVID-19 Pandemic* at 2 (Dec. 21, 2020), *available at* https://www.capolicylab.org/wp-content/uploads/2022/08/Dec-21st-Analysis-of-CA-UI-Claims-during-the-COVID-19-Pandemic.pdf.

[32] Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, 81 Proc. Machine Learning Rsch. 1, 12 (2018), *available at* https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf.

[33] Cal. Gov't Operations Agency, *Benefits and Risks of Generative Artificial Intelligence Report* at 14 (Nov. 2023), *available at* https://www.govops.ca.gov/wp-content/uploads/sites/11/2023/11/GenAI-EO-1-Report_FINAL.pdf.

[34] *Id.* at 21.

[35] The Government Operations Agency notes another similarity between GenAI systems and other AI systems built on neural networks, noting both may incorporate "black box algorithms that cannot provide direct explanations for their predictions." Indeed, GenAI systems and neural networks more generally have explainability and interpretability issues, which make it difficult to address cases where the technology produces an unexpected result that impacts the validity and consistency of the system's answers. Cal. Gov't Operations Agency, *Benefits and Risks of Generative Artificial Intelligence Report* at 4, 21 (Nov. 2023), *available at* https://www.govops.ca.gov/wp-content/uploads/sites/11/2023/11/GenAI-EO-1-Report_FINAL.pdf.

[36] Cal. Gov't Operations Agency, *Benefits and Risks of Generative Artificial Intelligence Report* at 23 (Nov. 2023), *available at* https://www.govops.ca.gov/wp-content/uploads/sites/11/2023/11/GenAI-EO-1-Report_FINAL.pdf.

Biased or otherwise poor quality training data can interfere with all types of AI systems and cause them to generate biased or discriminatory results.[37] The California Government Operations Agency report suggests that it is more difficult to resolve embedded bias in GenAI datasets, which are much larger than the datasets used to train many other automated systems.[38] Still, the report observed that government agencies must proactively assess these systems for algorithmic discrimination in high-impact areas to mitigate bias: a value statement that is applicable to all categories of AI systems.[39]

To reflect the State's commitment to creating "equitable outcomes" in the "deployment and implementation" of AI, the Guidelines should cover more than just GenAI.[40] As mentioned above, the Guidelines should apply when a state agency or department considers (1) "any automated system" that (2) has the potential to meaningfully impact the California public's "rights, opportunities, or access to critical resources or services."[41] Given the rapid pace of AI's technological development, a more inclusive and generally applicable definition would help ensure that California continues to foster a "safe and responsible innovation system that puts AI systems and tools to the best uses for Californians."[42]

## II. To Protect Vulnerable Communities, the Guidelines Should Require Community Input and a Focus on AI Uses That Impact Californians' Lives.

A clearer definition of AI will help agencies and departments understand when the Guidelines apply. For covered AI, the Executive Order directs that the Guidelines enable agencies and departments to consider the appropriateness of those systems for a given problem in light of its potential impact on vulnerable communities. Specifically, the Executive Order says the Guidelines should help agencies and departments decide

---

[37] *Id.*

[38] *Id.*; *see also* Timnit Gebru, Emily M. Bender, Angelina McMillan-Major, Margaret Mitchell, *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?* 🦜, in Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency 610 (2021), *available at* https://dl.acm.org/doi/10.1145/3442188.3445922.

[39] Cal. Gov't Operations Agency, *Benefits and Risks of Generative Artificial Intelligence Report* at 23 (Nov. 2023), *available at* https://www.govops.ca.gov/wp-content/uploads/sites/11/2023/11/GenAI-EO-1-Report_FINAL.pdf.

[40] Cal. Exec. Order N-12-23 §3(b) at 4 (Sep. 6, 2023), *available at* https://www.gov.ca.gov/wp-content/uploads/2023/09/AI-EO-No.12-_-GGN-Signed.pdf.

[41] *See* White House Off. of Sci. & Tech. Pol'y, *Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People* at 8 (Oct. 2022), *available at* https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf.

[42] Cal. Exec. Order N-12-23 §3(b) at 3 (Sep. 6, 2023), *available at* https://www.gov.ca.gov/wp-content/uploads/2023/09/AI-EO-No.12-_-GGN-Signed.pdf.

*whether* to deploy an AI system. [43] We propose the following preliminary process whenever an agency or department is considering deploying an AI system.

### A. The Guidelines Should Require Diverse Community Input to Identify the Problem and a Robust Examination of Non-AI Alternatives.

***Engage diverse communities to identify the problem.*** First, whenever AI might be considered, the Guidelines should instruct State agencies and departments to first identify the specific problem they want to solve, and to do this they should be required to consult with the people potentially impacted by the design, development, or use of the AI to discuss the problems they believe should be prioritized. This diverse engagement needs to include people impacted by the State program, other vulnerable populations, and workers impacted by agency use of AI.[44] The Guidelines should require that this engagement be robust and public, and that departments and agencies give significant weight to these voices in their identification of the problem to be addressed. With ample time prior to any decision of whether to deploy AI for a particular use, agencies and departments must disclose information about this feedback and the problem identification process.

***Identify both AI and non-AI possible solutions.*** Once the problem to be addressed is identified, the Guidelines should require that agencies and departments identify possible solutions – including those that do not involve AI as well those that do – to address the identified problems efficiently and without impacts on people's rights. This is important because AI will not necessarily be a solution to the specific or systemic problems facing Californians and their governments. [45]

***Assess possible solutions based on evidence.*** After an agency has identified the problem to be solved and the possible solutions, the Guidelines should require an evidence-based inquiry and public decision about whether an AI system is appropriate,

---

[43] The Executive Order asks for guidelines to guide decisions to "deploy" AI, decisions that could include both the deployment of an existing AI system for a new particular use and also a deployment that would require a new acquisition of AI. As such, the Guidelines should require the AIIA be conducted prior to any acquisition that would be necessitated by a deployment for a particular use. *See id.* at 4.

[44] As the White House Office of Science and Technology wrote, "[a]utomated systems should be developed with consultation from diverse communities, stakeholders, and domain experts to identify concerns, risks, and potential impacts of the system." *See* White House Off. of Sci. & Tech. Pol'y, *Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People* at 15 (Oct. 2022), *available at* https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf.

[45] As the Government Operations Agency has previously noted with regards to GenAI, it "may not always be optimal or necessarily appropriate for a state entity's programmatic needs or for the Californians they serve." The same is true for all forms of AI. Cal. Gov't Operations Agency et al., *State of California GenAI Guidelines for Public Sector Procurement, Uses and Training* at 7 (Mar. 2024), *available at* https://cdt.ca.gov/wp-content/uploads/2024/03/3a-GenAI-Guidelines.pdf.

informed by diverse input during an AI Impact Assessment (AIIA). This assessment, the basic components of which are described in Section III below, should determine *whether* to deploy AI for a particular use. Pursuant to the AIIA, the agency or department will utilize an evidence-based inquiry to determine the public benefits and potential public harms of the system. The department or agency should actively engage and seek input from affected community members, workers,[46] researchers, and civil rights experts, who can highlight potential benefits as well as bias, civil rights issues, privacy and security threats, and safety risks. This information should be made publicly available.

***Make a decision based on a public benefit and public harm standard.*** The Guidelines should state that AI systems may only be acquired or deployed in contexts that have consequences for people's lives if the department or agency, through such an evidence-based inquiry of the proposed use of AI, demonstrates a public benefit that substantially outweighs the potential public harm. Whether a proposed use of AI demonstrates a public benefit that substantially outweighs potential public harm should be determined by the overall AIIA process. The needs and input of potentially impacted communities should directly inform this decision.

* * *

AI systems should go through the AIIA process before being acquired or deployed. Though, crucially, in Section IV below, we also call on the Guidelines to recognize that some uses of AI are so harmful that they should not be considered for acquisition or deployment at all. The following subsections provide a framework and set of examples to help agencies and department when to conduct an AIIA prior to any decision to deploy an AI system.

### B. The Guidelines Should Require an AI Impact Assessment for a Broad Range of AI Uses.

The Guidelines should require that before any acquisition or deployment of AI in contexts that have consequences for people's lives, the department or agency should engage in a thoughtful consideration of the evidence-based benefits and risk of harm and complete an AI Impact Assessment (AIIA). Building upon a definition formulated by the U.S. Office of Management and Budget ("OMB"), and with the purpose of providing the flexibility for agencies and departments to account for emerging developments in the technology and fresh use cases of AI, the Guidelines should direct agencies and departments to use the AIIA process for any AI uses that may have consequences for people's lives include, but are not limited to:

[46] Agencies and departments should also involve employee groups and employees' union representatives in this process. *See infra* note 52, at 22 ("Consult and incorporate feedback from affected communities and the public").

AI whose output serves as a basis for a decision or action that has a legal, material, or other potential effect on an individual's or community's:

1. Civil rights, civil liberties, or privacy, including but not limited to freedom of speech, voting, human autonomy, and protections from discrimination, excessive punishment, and unlawful surveillance;

2. Equal opportunities, including equitable access to education, housing, insurance, credit, employment, and other programs where civil rights and equal opportunity protections apply; or

3. Access to or the ability to apply for critical government resources or services, including healthcare, financial services, public housing, social services, transportation, essential goods and services, and government benefits or privileges.[47]

These three broad categories can serve as a non-exhaustive set of criteria for agencies and departments to identify uses that are likely to impact people's rights and must conduct an AIIA prior to acquisition or deployment. Agencies and departments will then be able to draw on their own expertise about their missions to review AI use cases and determine whether they fall within the criteria.[48]

### C. The Guidelines Should Provide Examples of AI Uses That Require an AI Impact Assessment.

The Guidelines should also include an illustrative list of contexts and situations where the use of AI is likely to result in consequences for people's lives and where an AIIA impact assessment is necessary prior to acquisition or deployment. . The OMB's guidance for federal agencies in this respect can once again serve as a starting point that California can improve upon.[49] Agencies and departments should use this list to

---

[47] This definition draws from, but is not identical to, the definition of "rights-impacting AI" formulated by the U.S. Office of Management and Budget (OMB). While the OMB uses the term "rights-impacting," we do not use it in order to avoid any implication that agency or department decisions about the application of these guidelines are determinative of the scope of any person's rights. *See* Memorandum from Shalanda D. Young, Dir. of Off. of Mgmt. and Budget, on Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence at 29 (Mar. 28, 2024), *available at* https://www.whitehouse.gov/wp-content/uploads/2024/03/M-24-10-Advancing-Governance-Innovation-and-Risk-Management-for-Agency-Use-of-Artificial-Intelligence.pdf.

[48] As recognized by the California GenAI Guidelines for Public Sector Procurement, Uses and Training, state entity leaders will be responsible for decisions regarding the use of AI given each entity's "unique structure and mission." Cal. Gov't Operations Agency et al., *State of California GenAI Guidelines for Public Sector Procurement, Uses and Training* at 2 (March 2024), *available at* https://cdt.ca.gov/wp-content/uploads/2024/03/3a-GenAI-Guidelines.pdf.

[49] This bulleted list draws from, but is not identical to, one formulated by the Office of Management and Budget. Quotations featured in this list of presumptively rights-impacting uses are from that Memorandum. *See* Memorandum from Shalanda D. Young, Dir. of Off. of Mgmt. and Budget, on

quickly identify AI uses that can result in consequences for people's lives and necessitate thoughtful scrutiny through the AIIA process about whether they should even be acquired. The list should expand over time to account for developments and new understandings of AI risks, and at the start should include, but not be limited to:

- Decisions regarding speech, including "blocking, removing, hiding, or limiting the reach of protected speech;"
- "Replicating a person's likeness or voice without express consent;"
- Education-related decisions, specifically "detecting student cheating or plagiarism; influencing admissions processes; monitoring students online or in virtual-reality; projecting student progress or outcomes; recommending disciplinary interventions; determining access to educational resources or programs; determining eligibility for student aid or Federal education; or facilitating surveillance (whether online or in-person);"
- Housing-related decisions, specifically "[s]creening tenants; monitoring tenants in the context of public housing; providing valuations for homes; underwriting mortgages; or determining access to or terms of home insurance;"
- In employment, "[d]etermining the terms and conditions of employment, including pre-employment screening, reasonable accommodation, pay or promotion, performance management, hiring or termination, or recommending disciplinary action; performing time-on-task tracking; or conducting workplace surveillance or automated personnel management;"
- In healthcare, "[c]arrying out the medically relevant functions of medical devices; providing medical diagnoses; determining medical treatments; providing medical or insurance health-risk assessments; providing drug-addiction risk assessments or determining access to medication; conducting risk assessments for suicide or other violence; detecting or preventing mental-health issues; flagging patients for interventions; allocating care in the context of public insurance; or controlling health-insurance costs and underwriting;"
- In the financial realm, "[a]llocating loans; determining financial-system access; credit scoring; determining who is subject to a financial audit; making insurance determinations and risk assessments; determining interest rates; or determining financial penalties (e.g., garnishing wages or withholding tax returns);"
- Regarding government services or benefits, "[m]aking decisions regarding access to, eligibility for, or revocation of critical government resources or services; allowing or denying access…to IT systems for accessing services for benefits; detecting fraudulent use or attempted use of government services; assigning penalties in the context of government benefits;"

---

Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence at 33 (Mar. 28, 2024), *available at* https://www.whitehouse.gov/wp-content/uploads/2024/03/M-24-10-Advancing-Governance-Innovation-and-Risk-Management-for-Agency-Use-of-Artificial-Intelligence.pdf.

- Language services, including "[t]ranslating between languages for the purpose of official communication to an individual where the responses are legally binding; providing live language interpretation or translation, without a competent interpreter or translator present, for an interaction that directly informs an agency decision or action;"
- Uses of AI that entail the collecting, processing, and use of sensitive personal data, including children's data and data pertaining to an individual's race, sex, gender, ethnicity, religion, and national origin.[50]

## III. The Guidelines Should Require an AI Impact Assessment Prior to Decisions of Whether to Acquire or Deploy AI.

### A. Before the AI Impact Assessment

The Guidelines should require that agencies and departments conduct an AI Impact Assessment prior to any decision about whether to acquire or deploy a system. As described in Section II.A above, prior to making a decision to acquire or deploy AI for a use covered by the Guidelines, an agency or department should first ***engage diverse communities to identify the problem*** by soliciting the views of people who are potentially impacted by AI systems, including vulnerable communities, the people impacted by relevant state programs, and workers impacted by agency use of AI. Following that, the agency or department should ***identify both AI and non-AI possible solutions*** that could address the identified problems efficiently and without impacts on people's right and then ***assess possible solutions based on evidence*** using an AI impact assessment (AIIA). Finally, an agency or department should ***make a decision based on a public benefit and public harm standard***.

Basic components of this AIIA process are explained below. This is by no means comprehensive. This process is just one step towards ensuring "equitable outcomes in deployment and implementation of high-risk use cases" of AI that may impact vulnerable communities in California.[51]

### B. AI Impact Assessment

The AIIA should be a robust evidence-based analysis that helps agencies, departments, and the public understand the potential impact of the proposed AI use on the lives and civil rights of Californians and in particular vulnerable people. What

---

[50] Elec. Priv. Info. Ctr. (EPIC), *Comments of the Electronic Privacy Information Center to the Office of Management and Budget* at 10 (2023), *available at* https://epic.org/wp-content/uploads/2023/12/EPIC-OMB-AI-Guidance-Comments-120523-1.pdf. AI use cases that rely on these categories of data should be considered rights-impacting since they are "particularly vulnerable to inequitable or otherwise harmful outcomes," including data breaches and discrimination. *Id.*

[51] Cal. Exec. Order N-12-23 §3(b) at 4 (Sep. 6, 2023), *available at* https://www.gov.ca.gov/wp-content/uploads/2023/09/AI-EO-No.12-_-GGN-Signed.pdf.

follows is an illustrative (but not exhaustive) set of considerations, drawn from existing technology impact assessment frameworks and influenced by the OMB's guidance to federal agencies.[52] This assessment should include **(1)** an examination of the proposed purpose and evidence-based assessment of the benefit the AI system, **(2)** an explanation of how the AI system works and uses people's information, **(3)** an explanation of the potential harms of the AI system, and **(4)** how people will be people are involved and how Californians will be empowered with regards to the AI system. The Guidelines should direct the agency or department to publicly release the completed AIIA with significant time before a final decision about *whether* to move forward with the AI system is made.

To assist with this AI Assessment, departments and agencies should work with – and welcome feedback from – civil society experts and independent researchers with knowledge about civil rights, privacy, free speech, security, and discrimination law.[53] The Guidelines should prohibit agencies and departments from creating legal hurdles to efforts by experts and third-party researchers to assist in the evaluation of the risks of an AI system, including efforts to assess and provide feedback on the security, safety, or privacy of an AI system.

**The AI assessment should examine and document the specific problem, the purpose of the AI, and the demonstrated public benefits:**

- Based on conversations with potentially impacted people and the public process described in Section II.A above, the assessment should articulate the specific problem that the AI system will address, supported by specific metrics or qualitative analysis that demonstrate the AI system will address that problem. Metrics should be quantifiable measures of positive outcomes for the specified public purpose and the agency's mission that would be measurable.[54] Vendor claims or advertising about AI should not be taken at face value.[55] The analysis should explain how the AI

[52] The components of the assessment in Section III.B of this comment build on, but are not identical to, OMB's guidance to federal agencies. What we include here is a non-exhaustive list of elements, and we urge the team crafting the Guidelines to look at the entire OMB memo. Where applicable in this section, we include pincites to specific parts of that guidance. *See* Memorandum from Shalanda D. Young, Dir. of Off. of Mgmt. and Budget, on Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence at 33 (Mar. 28, 2024), *available at* https://www.whitehouse.gov/wp-content/uploads/2024/03/M-24-10-Advancing-Governance-Innovation-and-Risk-Management-for-Agency-Use-of-Artificial-Intelligence.pdf.

[53] *See supra* note 52, at 21 ("Identify and assess AI's impact on equity and fairness, and mitigate algorithmic discrimination when it is present").

[54] *See supra* note 52, at 17 ("The intended purpose for the AI and its expected benefit[s]"). In addition, a qualitative analysis can be used to demonstrate an expected positive outcome and that AI is better suited to accomplish the relevant task as compared to alternative strategies, including those that do not involve AI. *See supra* note 52, at 17 ("The intended purpose for the AI and its expected benefits"); *supra* note 52, at 18 ("Test the AI for performance in a real-world context").

[55] *See supra* note 52, at 19 ("Independently evaluate the AI").

system has been evaluated and demonstrated—including in other governmental contexts and in evaluations by researchers—to address the identified problem.[56]

**The AI assessment should explain how the AI system works and how it would use people's information:**

- The assessment should explain the information that has been used to train and evaluate the AI system as well as any information about Californians that the system may collect or use.[57] Agencies must assess the quality of the information used in the AI's design, development, training, testing, and operation and whether the agency can use that system in light of the AI's intended purpose. [58] An agency or department should not accept an AI vendor's assertion of trade secrets as a categorical bar to the government's ability to evaluate information about these systems.[59] The State should be particularly mindful of the serious privacy issues

---

[56] Agencies and departments should point to real world evidence, including testing, that demonstrates the AI system will work in its intended real-world context and produce known and measurable benefits for the identified public benefit. *See supra* note 52, at 18-19 ("Test the AI for performance in a real-world context").

[57] This is important because "[t]he harms that are consequences of the use of [large-scale language datasets and large-scale computer vision datasets] can stem from different points in the dataset creation process, from data collection, data annotation, data distribution, model training, model evaluation, and model inference." Researchers and experts Dr. Alex Hanna and Mehtab Khan offer a taxonomy for understanding the development of AI datasets and discuss why "[t]o clarify the scope of potential harms of large-scale AI datasets, it is necessary to identify different stakeholders impacted in the dataset development process." Khan, Mehtab and Hanna, Alex, *The Subjects and Stages of AI Dataset Development: A Framework for Dataset Accountability* 22, 28 (September 13, 2022). Forthcoming 19 Ohio St. Tech. L.J. (2023), *available at* SSRN: https://ssrn.com/abstract=4217148 or http://dx.doi.org/10.2139/ssrn.4217148.

[58] *Supra* note 52, at 18 ("The quality and appropriateness of the relevant data"). In conducting data assessments, the agency or department should request information about the system from the vendor. At a minimum, the agency or department should obtain sufficient descriptive information from the vendor (e.g., AI or data provider) that includes:

a. the purpose for which data subjects originally consented to the collection of their data;
b. the data collection and preparation process, which must also include the provenance of any data used to train, fine-tune, or operate the AI;
c. the quality and representativeness of the data for its intended purpose;
d. how the data is relevant to the task being automated and may reasonably be expected to be useful for the AI's development, testing, and operation;
e. whether the data contains sufficient breadth to address the range of real-world inputs the AI might encounter and how data gaps and shortcomings have been addressed either by the agency or vendor;
f. whether the data comes from an adequately reliable source; and
g. how errors in data entry, machine processing, or other sources are adequately measured and limited, to include errors from relying on AI-generated data as training data or model inputs.

[59] Indeed, it is standard regulatory practice to provide trade secrets in full to the government or a regulatory agency, and for that body (in coordination with the disclosing entity) to evaluate and disclose the information taking account of legitimate trade secret concerns. Numerous California

raised when an AI system is trained on the information of Californians that was originally collected for a different purpose.[60] The assessment should explain if information submitted by Californians will be used to further train the system.

- The assessment should include an explanation about whether the system will incorporate or be used to collect any information about protected classes as well as vulnerable or historically marginalized groups. Given the risks that can arise when AI correlates demographic information with other types of information, agencies should also assess and document whether the AI model could foreseeably use other attributes or information about a person as proxies and whether such use would address potential discrimination or equity concerns.[61]

**The AI assessment should explain the potential harms of the AI system, including the risk of discrimination and bias, threats to privacy and free speech, and other potential harms:**

- The assessment should document who will be potentially impacted by the proposed use of the system and any potential harms that may result from both the proper functioning of the system as well as from system failures, including failures in isolation and as a result of human users and other likely variables outside the scope of the system itself.[62] For Gen AI systems based on large language models, this includes explaining reliability and accuracy issues and their impact on the proposed purpose, including possible "hallucinations."[63] Agencies and departments should explain what, if anything, could be done to reduce and eliminate any negative impacts.
- The assessment should explain how the AI system would operate in a manner that would raise privacy or free speech issues. This includes whether and how use of the system will result in the collection, retention, or onward sharing of information about people who use the system, and whether and how the operation may generate

---

regulations in industries ranging from petroleum to toxic waste take this approach. *See e.g.*, Cal. Health & Saf. Code § 25257 (regarding chemicals in products, stating that "[t]his section does not authorize a person to refuse to disclose to the department information required to be submitted to the department pursuant to this article."); Cal. Health & Saf. Code § 25512(a)(3) (similar language regarding pollution and hazardous waste); Cal. Lab. Code § 7873(b)(1) (similar language regarding petroleum refineries).

[60] Article I, Section 1 of the California Constitution was also designed to "prevent[] government and business interests from collecting and stockpiling unnecessary information about us and from misusing information gathered for one purpose in order to serve other purposes or to embarrass us." *White v. Davis*, 533 P.2d 222, 233 (1974) (quoting ballot materials).

[61] *Supra* note 52, at 21 ("Identify and assess AI's impact on equity and fairness, and mitigate algorithmic discrimination when it is present").

[62] *See supra* note 52, at 17–18 ("The potential risks of using AI").

[63] *See* Gerrit De Vynck, *ChatGPT 'Hallucinates.' Some Researchers Worry It Isn't Fixable.*, Washington Post (May 31, 2023), *available at* https://www.washingtonpost.com/technology/2023/05/30/ai-chatbots-chatgpt-bard-trustworthy/.

connections or information about a person involving sensitive information, including but not limited to their identity, location, background, or associations.
- The assessment should include analyses for potential algorithmic discrimination, unlawful discrimination, harmful bias, or negative impacts on equity.[64]

**The AI assessment should also explain how people would be involved and how Californians would be empowered with regards to the proposed AI system:**

- The assessment should explain how people would be meaningfully informed about the proposed use of an AI system and any way that it may impact them. To be clear, notice is no substitute for preventing harm or protecting people's rights. But people deserve to know when, why, and how a government agency is considering using an AI system as any part of a decision that concerns them.[65]
- The assessment should explain what the role of humans would be in the proposed use of AI. They should explain how humans with the authority to exercise oversight, and intervention will be part of decisions or actions that involve AI and could result in significant impacts to people's lives. As part of this, agencies and departments should identify situations where AI functionality will not be permitted to intervene directly or make a decision without human involvement, consideration, and accountability.[66] The agency should also explain if and how the agency would provide and maintain a mechanism by which a person will be able to opt out of the proposed use of an AI system in favor of a human alternative, and how that opt out will avoid imposing burdens on access to a government service.
- The assessment should explain how a person will appeal, obtain human review, or otherwise contest an AI system's decision about them.[67] As part of this, the department or agency should explain how this process leverages or expands existing processes that already exist for non-AI systems. Whatever their form, these processes should be easy to use and not overly burdensome.

---

[64] *See supra* note 52, at 21 ("Identify and assess AI's impact on equity and fairness, and mitigate algorithmic discrimination when it is present").
[65] *See supra* note 52, at 21 ("Provide public notice and plain-language documentation"). Though this comment focuses on the decision of whether to deploy AI, the agency or department can also explain if and how the agency would provide and maintain a mechanism by which a person will be able to opt out of the proposed use of an AI system in favor of a human alternative, and how that opt out will avoid imposing burdens on access to a government service. *See* supra note 88, at 24 ("Maintain options to opt-out for AI-enabled decisions").
[66] *See supra* note 52, at 23 ("Maintain human consideration and remedy processes"); *id.* at 20 ("Provide additional human oversight, intervention, and accountability as part of decisions or actions that could result in a significant impact on rights or safety").
[67] *See supra* note 52, at 23 ("Maintain human consideration and remedy processes"). Any such method of recourse shall be optional for a person and would be in addition to a person's ability to bring a suit in state or federal court to, among other things, vindicate their rights or contest the decision.

### C. Apply the Public Benefit Standard to Proposed and Existing AI Systems.

Using the AI impact assessment and the input and preferences of potentially impacted communities, the agency or department should apply the public benefit standard, assessing whether such a purpose for the AI has been demonstrated and where the evidence-based benefits of the particular use of AI substantially outweigh the potential harms. As stated above, the answer of whether there is a public benefit that substantially outweighs potential public harms or not should be determined by conversations with potentially impacted communities and the overall AIIA process.

It is important that agencies and departments also apply these Guidelines to reexamine existing AI systems to ensure those systems are helping and not harming Californians. Thus, the Guidelines should require that any agency or department that already uses an AI system follow the same steps: incorporating the feedback of impacted communities, conducting an AIIA, and applying the public benefit standard. As with proposed new uses of AI, this AIIA should be publicly released well before any decision is made.

Though this comment is focused on the process and assessment of AI systems prior to decisions about whether to deploy them, it is important that there be regular, independent public evaluation for any AI systems that are actually deployed.[68] These evaluations should be more than technical. They should evaluate the overall efficacy of the system and analyze whether the system is achieving public benefits for its stated purpose and any public harms related to the system and its impact on people, particularly vulnerable people.

In addition to the AIIA above, the Guidelines should require that an agency or department confirm that any use of an existing AI system follows the principle of data minimization and that any collection of information about people will be limited to what is necessary to achieve the public benefit for the stated purpose.[69] Furthermore, because some vendors may seek to use Californians' information to further train AI

---

[68] This evaluation should be independent and "must also include oversight and consideration by an appropriate internal agency authority not directly involved in the system's development or operation." *See supra* note 52, at 20 ("Regularly evaluate risks from the use of AI"). The OMB memo discusses a number of ways to evaluate AI systems for effectiveness, bias, and discrimination, including on an ongoing basis. *See, e.g., Id.* at 19 ("Regularly evaluate risks from the use of AI"); id. ("Conduct ongoing monitoring"); *id*. at 20 ("Mitigate emerging risks to rights and safety"); *id*. at 23 ("Conduct ongoing monitoring and mitigation for AI-enabled discrimination").

[69] Data minimization means only collecting, using, retaining, or sharing information that is necessary to serve the specified purpose. As part of data minimization, an agency or department should explain how they will ensure information will not be retained for longer than necessary—by either the agency or the vendor —to achieve the identified purpose. *See* Data & Soc'y, *Response to the Request for Comments on OMB's Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence Draft Memorandum* 14, 15 (2023), *available at* https://datasociety.net/wp-content/uploads/2023/12/DS-Comment-on-OMB-memo-12-5-23.pdf.

systems without their knowledge or consent, the Guidelines should require a department or agency confirm via agreement that any AI vendor only uses information provided to it to serve the public benefit for the stated purpose for that system and the particular use authorized by the person whose information is at issue.

The Guidelines should specify that whenever an existing use of an AI system cannot, based on the input of impacted communities and the AI assessment, be justified with a public benefit and where the evidence-based benefits of the use of the AI system outweigh its potential harms, the agency or department should decommission that use. Agencies and departments must discontinue use of the AI system in a safe manner.[70]

## IV. The Guidelines Should Identify AI Uses That Are Too Harmful to Pursue.

Finally, the Guidelines should recognize that some uses of AI are too harmful to be deployed by State agencies under any circumstances. Accordingly, the Guidelines should prohibit agency procurement and use of technologies that pose an unacceptable risk to Californians. Specifically, the Guidelines and your office should prohibit agencies from using AI systems that include **(A)** facial recognition or other biometric surveillance systems, **(B)** predictive policing systems, **(C)** emotion detection systems, **(D)** family policing systems, and **(E)** criminal justice systems.

We are only at the beginning of the AI age. To ensure this list can be expanded in the future, the Guidelines should set forth a process by which this list can be reassessed and expanded to implement prohibitions of other high-risk AI systems.

### A. Facial recognition and other biometric surveillance systems.

The Guidelines should prohibit agencies from using facial and biometric surveillance systems. Face surveillance is dangerous both because its inaccuracies can lead to harmful misidentifications and because it facilitates pervasive surveillance. Today, we know of seven wrongful arrest cases resulting from incorrect facial recognition results.[71] In almost every instance, the person falsely matched using face recognition was Black.[72] Testing by the National Institute for Standards & Technology

---

[70] *See supra* note 52, at 20 ("Where the AI's risks to rights or safety exceed an acceptable level and where mitigation strategies do not sufficiently reduce risk, agencies must stop using the AI as soon as is practicable.")

[71] Nathan F. Wessler, *Why aren't California lawmakers banning police from using facial recognition technology?*, Sacramento Bee (Apr. 25, 2024), *available at* https://www.sacbee.com/opinion/op-ed/article287750520.html; *see also* ACLU, *Comment on Law Enforcement Agencies' Use of Facial Recognition Technology, Other Technologies Using Biometric Information, and Predictive Algorithms (Executive Order 14074, Section 13(e))* at 4 (Jan. 19, 2024), *available at* https://www.aclu.org/documents/aclu-comment-facial-recognition-and-biometric-technologies-eo-14074-13e.

[72] ACLU, *Comment on Law Enforcement Agencies' Use of Facial Recognition Technology, Other Technologies Using Biometric Information, and Predictive Algorithms (Executive Order 14074,*

(NIST) has found that facial recognition algorithms were up to 100 times more likely to misidentify Asian and African American people than white men, and that women and younger individuals were also subject to disparately high misidentification rates.[73] Efforts to set numerical thresholds[74] or require additional confirmatory evidence[75] do not prevent police from acting on flawed outputs. Beyond these concerns with accuracy, deployment of face surveillance for real-time tracking through video feeds would give the government an unprecedented power to track people's daily movements and exercise of First Amendment activities. This would pose a serious threat to privacy, free speech, and the right to travel. Recognizing these harms, five California localities have already stopped law enforcement's use of face surveillance.[76] The Guidelines should not permit State agencies or departments to use this kind of surveillance system.

Biometric surveillance systems that exploit other parts of the body to enable the secretive monitoring of Californians should also be prohibited. In 2015, researchers at Carnegie Mellon University demonstrated iris recognition scanners that could capture and identify irises from up to 40 feet away.[77] The researchers demonstrated this technology through a simulated traffic stop scenario; deploying it in an actual traffic stop would be cause for serious concern. The use of iris scans on people in this way is invasive and would intrude upon people's reasonable expectations that their eyes will not be the basis for identification or comprehensive tracking.

### B. Predictive policing systems.

The Guidelines should also prohibit agencies from using predictive policing technology. Predictive policing software typically utilizes historical crime data to generate predictions about suspected future crime and criminality, such as where and when it is likely to take place, or even which specific people are more likely to commit

---

*Section 13(e))* at 4 (Jan. 19, 2024), *available at* https://www.aclu.org/documents/aclu-comment-facial-recognition-and-biometric-technologies-eo-14074-13e; *see also* Complaint, *Williams v. City of Detroit*, 2:21-cv-10827 (E.D. Mich. 2021).

[73] Patrick Grother, Mei Ngan, Kayee Hanaoka, U.S. Dep't of Com., Nat'l Inst. for Standards & Tech., *Face Recognition Vendor Test Part 3: Demographic Effects* 2–3, 8 (Dec. 2019), *available at* https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf.

[74] Matt Cagle and Marissa Gerchick, *When it Comes to Facial Recognition, There is No Such Thing as a Magic Number*, ACLU of NorCal (Feb. 7, 2024), *available at* https://www.aclunc.org/blog/when-it-comes-facial-recognition-there-no-such-thing-magic-number.

[75] Indeed, "in most of the known cases of face recognition wrongful arrests… police did try to confirm the match, but then arrested the wrong person anyway." Nathan F. Wessler, *Why aren't California lawmakers banning police from using facial recognition technology?*, Sacramento Bee (Apr. 25, 2024), *available at* https://www.sacbee.com/opinion/op-ed/article287750520.html.

[76] *The Fight Against Surveillance in* San Francisco, ACLU of NorCal, *available at* https://www.aclunc.org/campaign/fight-against-surveillance-san-francisco; *The Fight to Stop Face Recognition Technology*, ACLU (last updated Jun. 7, 2023), *available at* https://www.aclu.org/news/topic/stopping-face-recognition-surveillance.

[77] *See* Brooks Hays, *Iris Scanner Can ID a Person from 40 Feet Away*, UPI (May 22, 2015), *available at* https://www.upi.com/Science_News/2015/05/22/Iris-scanner-can-ID-a-person-from-40-feet-away/7071432303037.

crimes.[78] However, historical crime data is incomplete, racially skewed, and reflects centuries of ineffective and biased policing practices.[79] Relying on such data to predict future crime perpetuates the profiling of individuals and communities, facilitates needless interactions between police and people that can turn dangerous, and erodes trust between police and the public.[80] Studies of the software in practice suggest that it does not predict crime effectively.[81] One investigative piece examining 23,631 predictions generated by one such software product determined that it had a success rate of less than one-half of one percent: "Fewer than 100 of the predictions lined up with a crime in the predicted category, that was also later reported to police."[82] Here in California, when researchers compared public health survey data about illegal drug use in Oakland against police records of drug arrests in Oakland, they demonstrated that drug arrests were concentrated in primarily low-income non-white neighborhoods, while drug use was more widely distributed.[83] Researchers then fed their data into the predictive algorithm published by the company PredPol. They found that "rather than correcting for the apparent biases in the police data, the model reinforces these biases."[84]

### C. Emotion detection systems.

The Guidelines should prohibit State agencies and departments from using AI systems that purport to detect emotions.[85] Some of these systems operate by observing people's faces. These systems cannot live up to their claim of accurately detecting and classifying people's emotions because, as one recent systematic review of the existing scientific evidence concluded, the idea that people reliably express and recognize

---

[78] Ezekiel Edwards, *Predictive Policing Software Is More Accurate at Predicting Policing Than Predicting Crime*, ACLU (Aug. 31, 2016), *available at* https://www.aclu.org/news/criminal-law-reform/predictive-policing-software-more-accurate.

[79] Rashida Richardson, Jason M. Schultz, Kate Crawford, *Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice*, New York University Law Review Online (2019), *available at* https://www.nyulawreview.org/online-features/dirty-data-bad-predictions-how-civil-rights-violations-impact-police-data-predictive-policing-systems-and-justice/.

[80] *Id.*; Ezekiel Edwards, *Predictive Policing Software Is More Accurate at Predicting Policing Than Predicting Crime*, ACLU (Aug. 31, 2016), *available at* https://www.aclu.org/news/criminal-law-reform/predictive-policing-software-more-accurate.

[81] Aaron Sankin and Surya Mattu, *Predictive Policing Software Terrible At Predicting Crimes*, Markup (Oct. 2, 2023), *available at* https://themarkup.org/prediction-bias/2023/10/02/predictive-policing-software-terrible-at-predicting-crimes.

[82] *Id.*

[83] Kristian Lum and William Isaac, *To predict and serve?*, 13 Significance 14, 15–16 (2016), *available at* https://perma.cc/BX9F-YLKM.

[84] Kristian Lum and William Isaac, *To predict and serve?*, 13 Significance 14, 18 (2016), *available at* https://perma.cc/BX9F-YLKM.

[85] Notably, the Government Operations Report on "Benefits and Risks" of GenAI flags as a possible use case the employment of GenAI to conduct "sentiment analysis of public feedback on state policies." Cal. Gov't Operations Agency, *Benefits and Risks of Generative Artificial Intelligence Report* at 7–12 (Nov. 2023), *available at* https://www.govops.ca.gov/wp-content/uploads/sites/11/2023/11/GenAI-EO-1-Report_FINAL.pdf.

certain emotions in specific facial expressions has insufficient evidence to support it.[86] In fact, "very little about how and why certain facial movements express instances of emotion is actually known at a level of detail that such conclusions could be used in important, real-world applications."[87]

Other purported emotion recognition systems rely on voices. These systems are also flawed. First, existing AI tools that convert speech-to-text can be significantly inaccurate. One study found automated speech recognition products were struggled to properly recognize common agreement and disagreement sounds like "Mm-hm" and "Uh-uh."[88] A recent study of one prominent system, OpenAI's Whisper, found that approximately 1% of its audio transcriptions contained hallucinated phrases or sentences.[89] The authors concluded that "38% of hallucinations include explicit harms such as violence, made up personal information, or false video-based authority."[90] If transcribed text is not accurate, then any conclusions about its emotional content may also be based on falsity. In addition to inaccuracy, purported emotion recognition systems demonstrate significant bias along racial and gender lines: employability language models found that "stronger associations with [African American English] correlate with lower occupational prestige."[91] Used in a number of contexts that consequences for people's lives—from administration of services or benefits to hiring of workers—these systems demonstrate a propensity to negatively impact marginalized communities. The lack of consensus on theoretical frameworks and practical indicators of emotion result in significant ethical considerations against utilizing AI systems reliant on them.[92] Emotion predictors in AI are inseparable from their origins in

---

[86] Lisa F. Barrett, Ralph Adolphs, Stacy Marsella, Aleix Martinez, Seth D. Pollak, *Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movement*, 20 Psych. Sci. Pub. Interest 48 (2019), *available at* https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6640856/pdf/nihms-1021596.pdf; *see also* Jay Stanley, *Experts Say 'Emotion Recognition' Lacks Scientific Foundation*, ACLU (Jul. 18, 2019), *available at* https://www.aclu.org/news/privacy-technology/experts-say-emotion-recognition-lacks-scientific.

[87] *Id.* at 51.

[88] Brian D Tran, Kareem Latif, Tera L Reynolds, Jihyun Park, Jennifer Elston Lafata, Ming Tai-Seale, Kai Zheng, "*Mm-hm," "Uh-uh": are non-lexical conversational sounds deal breakers for the ambient clinical documentation technology?*, Journal of the American Medical Informatics Association, Volume 30, Issue 4, April 2023, Pages 703–711, https://doi.org/10.1093/jamia/ocad001.

[89] Allison Koenecke, Anna S. G. Choi, Katelyn Mei, Hilke Schellmann, Mona Sloane, *Careless Whisper: Speech-to-Text Hallucination Harms* (Feb. 2024), *available at* https://arxiv.org/pdf/2402.08021.pdf.

[90] Allison Koenecke, Anna S. G. Choi, Katelyn Mei, Hilke Schellmann, Mona Sloane, *Careless Whisper: Speech-to-Text Hallucination Harms* (Feb. 2024), *available at* https://arxiv.org/pdf/2402.08021.pdf.

[91] Valentin Hofmann, Pratyusha R. Kalluri, Dan Jurafsky, Sharese King, *Dialect Prejudice Predicts AI Decisions about Character, Employability, and Criminality* at 37 (2024), *available at* https://arxiv.org/pdf/2403.00742.pdf.

[92] Luke Stark and Jesse Hoey, *The Ethics of Emotion in Artificial Intelligence Systems* at 786 (Mar. 2021), *available at* https://dl.acm.org/doi/pdf/10.1145/3442188.3445939.

normative Western philosophy and science,[93] producing biased and inaccurate conclusions.

Even if emotion recognition technology were capable of accurate conclusions, California agencies should not use AI systems to make decisions about people based on their emotional state. People must be free to feel, show, and communicate emotions without fear that their government will use automated systems to judge them or make decisions about them based on those emotions. California residents should not have to live in fear that AI systems will be making decisions about their lives based on their emotional state.

### D. Criminal justice decisions.

As things stand, we are deeply concerned about the use of AI systems to make decisions in the context of the criminal justice system. Already, a number of automated decision-making systems exist that purport to help agencies make determinations related to sentencing, parole, supervised release, probation, bail, pretrial release, or pretrial detention. Such systems raise serious questions of not only relating to *whether* they work, but also *how* they obscure complex policy decisions and judgment calls behind a veil of objectivity.[94] In California, police tested an "intelligence tool" that assigned "threat levels" to residents based on possibly flawed and inaccurate information.[95] Elsewhere, multiple reports have shown that gunshot detection systems are flawed in methodology and effectiveness.[96] Regardless of accuracy, folding AI systems in the criminal justice context risks obscuring significant policy judgments—for example, the federal government reportedly secretly altered the threshold for what would be considered "minimum risk" under the PATTERN Risk Assessment Tool,

---

[93] Abeba Birhane, *Automating Ambiguity: Challenges and Pitfalls of Artificial Intelligence* at 32 (Oct. 2021) (Ph.D. dissertation, University of Dublin), *available at* https://arxiv.org/pdf/2206.04179.pdf%20page%2032.pdf; *See also* Ifeoma Ajunwa, *Automated Video Interviewing as the New Phrenology* 36 Berkeley Tech. L. J. 1174 (2021), *available at* https://btlj.org/wp-content/uploads/2023/01/0008-36-3-Ajunwa_Web.pdf.

[94] *See* Arvind Narayanan, *How to Recognize AI Snake Oil*, Princeton University, *available at* https://www.cs.princeton.edu/~arvindn/talks/MIT-STS-AI-snakeoil.pdf (describing how AI has not been meaningfully demonstrated to be capable of predicting social outcomes, including outcomes across an array of criminal justice contexts).

[95] Matt Cagle, *This Surveillance Software is Probably Spying on #BlackLivesMatter*, ACLU of NorCal (Dec. 15, 2015), *available at* https://www.aclunc.org/blog/surveillance-software-probably-spying-blacklivesmatter.

[96] Russell Contreras, *Critics say gunshot-detection technology often doesn't work*, Axios (Apr. 9, 2022), *available at* https://www.axios.com/2022/04/07/%20campaign-zero-against-shotspotter-crime; Garance Burke, Martha Mendoza, Juliet Linderman, Michael Tarm, *How AI-powered tech landed man in jail with scant evidence*, AP News (Mar. 5, 2022), *available at* https://apnews.com/article/artificial-intelligence-algorithm-technology-police-crime-7e3345485aa668c97606d4b54f9b6220.

meaning far fewer people would qualify to be considered for release.[97] Issues like these seriously call into question the use of AI systems to make decisions that impact people's liberty and lives in the criminal context.

### E. Family policing systems.

Finally, the Guidelines should prohibit the use of AI systems for the purpose of family-policing.[98] Many child welfare agencies have begun turning to automated risk assessment systems for reasons ranging from trying to predict which children are at higher risk for maltreatment to improving agency operations.[99] In 2021, the ACLU released a report entitled Family Surveillance by Algorithm surveying the use of these systems in counties and states around the country, including in California.[100] Despite their growing use by state and local governments, few families or advocates have heard about them, much less provided meaningful input into their development and use.

Research on family policing systems has demonstrated serious civil rights concerns. For example, Carnegie Mellon University researchers examined a "family screening tool" deployed in Allegany County, Pennsylvania, and they found that the software flagged a disproportionate number of Black children for mandatory investigation.[101] Their research suggests that algorithmic screening programs used to

---

[97] Marissa Gerchick, Brandon Buskey, Aaron Horowitz, Malika Mohan, Suresh Venkatasubramanian, Kweku Kwegyir-Aggrey, *Don't Let the Math Distract You: Together, We Can Fight Algorithmic Injustice*, ACLU (Aug. 8, 2023), *available at* https://www.aclu.org/news/criminal-law-reform/we-can-fight-algorithmic-injustice.

[98] As defined by the UpEnd Movement, the term "family policing system" is used "to refer to what has more commonly been known as the child welfare system. We believe this term more accurately captures the roles this system plays in the lives of families, which include surveillance, regulation, and punishment, all roles associated with policing rather than children's welfare. These roles are used to maintain the control and oppression of Black, Native, and Latinx families, which is also consistent with the practice of policing." *Fighting to End the Family Policing* System, ACLU of NorCal, *available at* https://www.aclunc.org/issue/gender-sexuality-reproductive-justice/fighting-end-family-policing-system. In the AI context, this may include systems that "[p]rovid[e] recommendations, decisions, or risk assessments about adoption matching, child protective actions, recommending child custody, whether a parent or guardian is suitable to gain or retain custody of a child, or protective actions for senior citizens or disabled persons." Memorandum from Shalanda D. Young, Dir. of Off. of Mgmt. and Budget, on Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence at 33 (Mar. 28, 2024), *available at* https://www.whitehouse.gov/wp-content/uploads/2024/03/M-24-10-Advancing-Governance-Innovation-and-Risk-Management-for-Agency-Use-of-Artificial-Intelligence.pdf.

[99] Christopher Teixeira and Matthew Boyas, MITRE Corp., *Predictive Analytics in Child Welfare: An Assessment of Current Efforts, Challenges and Opportunities* at 6–7 (2017), *available at* https://perma.cc/5PFR-BCV9.

[100] Anjana Samant, Aaron Horowitz, Kath Xu, Sophie Beiers, *Family Surveillance by Algorithm: The Rapidly Spreading Tools Few Have Heard Of*¸ ACLU (Sep. 29, 2021), *available at* https://www.aclu.org/documents/family-surveillance-algorithm.

[101] Logan Stapleton et al., *Extended Analysis of 'How Child Welfare Workers Reduce Racial Disparities in Algorithmic Decisions'* at 3 (Apr. 29, 2022) *available at* https://loganstapleton.com/wp-

flag instances of child neglect can exacerbate racial disparities.[102] California has already tried using predictive analytics programs for family policing, with deeply troubling results. In 2019, the California Department of Social Services shut down a three-year-old program that relied on such a tool after concluding that it "would not keep kids safer, and could lead to racial profiling," finding that "[a]bout 90% of cases ranked as 'low risk' by the [model] had safety threats" and that because the system relies on biased past actions, the harms would disproportionately be to Black and Native American families "who have historically been profiled as higher risk."[103]

The fact is that countless policy choices and value judgments are made in the course of creating and using risk-assessment systems like those deployed in the family policing system, and those choices can impact whether the tool promotes "fairness" or reduces racial disproportionality in families and communities. Given the disastrous consequences of these systems—disproportionately wrenching apart women and children who are Indigenous[104], Black,[105] or experiencing poverty[106]—the use of automated tools as part of the child-welfare system should be prohibited.

## V. Your Office Should Assist Localities Seeking to Apply the Guidelines.

Although the Guidelines are aimed at helping California "state agencies and departments,"[107] many local uses of automated decision-making technology are already impacting vulnerable Californians. Your office should explore providing assistance to localities that wish to adhere to the Guidelines voluntarily. This includes extending available resources, including technical assistance and training, to aid these localities who are already using AI systems or who may consider deploying it.

A number of California localities already use AI. As discussed above, local law enforcement agencies have deployed predictive policing software. Among many other

---

content/uploads/2022/04/Extended_Analysis__How_Child_Welfare_Workers_Reduced_Racial_Disparities_in_Algorithmic_Decisions.pdf ("Carnegie Mellon Study").

[102] *Id.* at 2.

[103] Jeremy Loudenback, *The Foster Care System Turns to Big Data: Promising or Profiling?* (Feb. 1, 2022), *available at* https://imprintnews.org/child-welfare-2/the-foster-care-system-turns-to-big-data-promising-or-profiling/62359; *see also* Sallo Ho and Garance Burke, *Here's how an AI tool may flag parents with disabilities*, AP News (Mar. 15, 2023), *available at* https://apnews.com/article/child-protective-services-algorithms-artificial-intelligence-disability-02469a9ad3ed3e9a31ddae68838bc76e.

[104] Nat'l Indian Child Welfare Ass'n and Pew, *Time for Reform: A Matter of Justice for American Indian and Alaskan Native Children* (2007), *available at* https://www.pewtrusts.org/-/media/legacy/uploadedfiles/wwwpewtrustsorg/reports/foster_care_reform/nicwareportpdf.pdf.

[105] Alan J. Dettlaff and Reiko Boyd, *Racial Disproportionality and Disparities in the Child Welfare System: Why Do They Exist, and What Can Be Done to Address Them?*, 692 Annals Am. Acad. Pol. & Soc. Sci. 253 (2020), *available at* https://journals.sagepub.com/doi/abs/10.1177/0002716220980329.

[106] Jerry Milner and David Kelly, *It's Time to Stop Confusing Poverty With Neglect*, The Imprint (Jan. 17, 2020), *available at* https://imprintnews.org/child-wel- fare-2/time-for-child-welfare-system-to-stop-confusing-poverty-with-neglect/40222.

[107] Cal. Exec. Order N-12-23 §3(b) at 3 (Sep. 6, 2023), *available at* https://www.gov.ca.gov/wp-content/uploads/2023/09/AI-EO-No.12-_-GGN-Signed.pdf.

examples, the City of Los Angeles has used an algorithm-based system to allocate scarce affordable housing to the unhoused.[108] But the system created stark racial disparities in housing by favoring White adults over Black and Latino adults in their access to housing.[109] If impacted communities are not consulted and decisions to deploy are not scrutinized, localities will continue to deploy AI systems that do more harm than good.

The Guidelines, and the expertise the California state government is developing in AI, could be of considerable use to localities as they consider whether to integrate AI into their decision-making and, if so, how to do so responsibly. One way to do this is to offer technical assistance and expertise to localities seeking to implement the Guidelines. This could be a relatively light lift, such as opening up already-planned webinars and other trainings to local officials, and making implementation toolkits available for the Guidelines, as it has already done for its earlier guidance under 3(a).

Thank you for considering our comments. We look forward to continued conversations to ensure that California can thoughtfully consider AI technology as it seeks to address longstanding systemic challenges and make life better for vulnerable Californians.

Sincerely,

Matt Cagle
Senior Staff Attorney
Technology & Civil Liberties Program
ACLU of Northern California

Nicole Ozer
Technology & Civil Liberties Director
Technology & Civil Liberties Program
ACLU of Northern California

Carmen-Nicole Cox
Director of Government Affairs
ACLU California Action

[108] Colin Lecher and Maddy Varner, *L.A.'s Scoring System for Subsidized Housing Gives Black and Latino People Experiencing Homelessness Lower Priority Scores*, Markup (Feb. 28, 2023), *available at* https://themarkup.org/investigation/2023/02/28/l-a-s-scoring-system-for-subsidized-housing-gives-black-and-latino-people-experiencing-homelessness-lower-priority-scores.
[109] *Id.*