CARNEGIE
ENDOWMENT FOR
INTERNATIONAL PEACE

OCTOBER 2025

Carnegie California AI Survey

Ian Klaus, Mark Baldassare, Rachel George,
Scott Kohler, Marissa Jordan, Abigail Manalese

Carnegie California AI Survey

Ian Klaus, Mark Baldassare, Rachel Ann George,
Scott Kohler, Marissa Jordan, Abigail Manalese

Carnegie Endowment for International Peace
Publications Department
1779 Massachusetts Avenue NW
Washington, DC 20036
P: + 1 202 483 7600
F: + 1 202 483 1840
CarnegieEndowment.org

This publication can be downloaded at no cost at CarnegieEndowment.org.

Contents

Summary

The 2025 Carnegie California AI Survey offers the broadest statewide survey to date on artificial intelligence. Though the technology is evolving at a rapid rate and its impacts on democracy and the economy remain uncertain, it has already moved to the center of policymaking debates. California is home not only to the industry leaders driving the technological change but also to the policy innovations that might shape its future, including the Transparency in Frontier Artificial Intelligence Act, signed into law in late September 2025.

While Carnegie California surveyed Californians on AI in 2023 and 2024 in the context of global affairs, the 2025 Carnegie California AI Survey offers deeper insight into how the state's residents think about the technology in terms of their work, privacy, safety, and communities, as well as its impacts on the economy and the nation's democracy.

Key findings, captured below, reveal anxiety and uncertainty around AI impacts. They also reveal notable gender and geographic divides, but, importantly, in the state and national context, areas of substantial alignment between Republicans and Democrats.

- Most Californians say that AI is important to the economic growth and competitiveness of both the United States and California. These views are held across political affiliations. There is anxiety, however, that the technology will have negative impacts on employment opportunities and deepen inequality in the near term.

- Most employed Californians believe AI is being used in their workplaces, but they are unlikely to have received training in or altered their career choice because of the technology. Less than half of employed Californians are using AI on a daily basis

or occasionally at work, while about half expect to be using AI more in the future at work. A majority of employed residents say they are interested in being offered training and courses on AI uses, and three in four think that skills to understand and use AI will be important for a worker to be successful in today's economy.

- Californians do not believe AI has yet improved government processes or service delivery, highlighting gaps in its use and/or communication about its potential use across the state. They also hold mixed opinions about whether it should be deployed to do so, though women are less supportive than men of government use of AI. Unease about government use of AI spans the political spectrum.

- Most Californians report broad concern over a number of AI-related risks, with privacy, spam, misinformation, and cybersecurity as especially prevalent. Significant majorities support public policies to temper these risks, such as protection for whistleblowers and workers whose jobs are threatened with displacement. Likewise, overall majorities believe that safety should be prioritized over innovation, though there are some partisan differences, with Democrats more inclined than Republicans to prioritize safety. As it comes to developing guardrails for safety, Californians support a nationwide effort across civil society, industry, government, and universities, showing only a small partisan divide. They believe industry should be involved in helping establish such guidelines and that government should require AI companies to test their most advanced systems for safety and provide a detailed plan for how they will prevent harm.

Methodology

This is the first year of the Carnegie California AI Survey. Findings in this paper are based on a survey of 1,601 adult residents of California, including an oversample of 101 California adults who attended community college. The survey was conducted via YouGov between June 27 and July 24, 2025, in English and Spanish according to respondents' preferences. The questions in four topic areas were designed by the Carnegie California survey team. The Carnegie California survey team invited input, comments, and suggestions from policy experts and its own advisory group—including advisers from state and local government, California universities and think tanks, industry, and civil society—during workshops in December 2024 and January 2025. However, survey methods, questions, and content were solely determined by the Carnegie California survey team.

YouGov fielded two separate surveys for this project. The first survey interviewed 208 Californian adults, with the goal of screening for and surveying 100 current or former community college students. The 208 were then matched down to a sample of 200, yielding a target subsample of 101. The respondents were matched to a sampling frame on

gender, age, race, and education. The sampling frame demographics are based on the 2023 American Community Survey (ACS) public use microdata file. The matched cases were weighted to the sampling frame using propensity scores. The matched cases and the frame were combined, and a logistic regression was estimated for inclusion in the frame. The propensity score function included age, gender, race/ethnicity, and years of education. The propensity scores were grouped into deciles of the estimated propensity score in the frame and post-stratified according to these deciles. The weights were then post-stratified on 2024 presidential vote choice and ranked along a four-way stratification of gender, age, race, and education. The weighted dataset of 200 was then subsetted on the 101 California adults who are current or former community college students, and the weights were trimmed and recentered around 1, to produce the final oversample weights. In the second survey, YouGov interviewed 1,510 Californian adults who were then matched down to a sample of 1,500. The respondents were matched to a sampling frame on gender, age, race, and education. The sampling frame demographics are based on the 2023 American Community Survey (ACS) public use microdata file.

The final dataset from the oversample data (n=101) was merged with matched data to produce the final dataset of 1,601 California residents aged 18 or older. The merged cases were weighted to the sampling frame using propensity scores. The merged cases and the frame were combined, and a logistic regression was estimated for inclusion in the frame. The propensity score function included age, gender, race/ethnicity, and years of education. The propensity scores were grouped into deciles of the estimated propensity score in the frame and post-stratified according to these deciles. The weights were then post-stratified on 2024 presidential vote choice and raked along a four-way stratification of gender, age, race, and education.

The YouGov panel includes information about each respondent's demographic and political profile, used in this paper. We present results for four racial/ethnic groups: Asian, Black, Hispanic, and White. Residents of other racial and ethnic groups are included in the results reported for all adults, but sample sizes for these less populous groups are not large enough to report separately. We present results for five geographic regions, accounting for approximately 90 percent of the state population. "Central Valley" includes the counties Butte, Colusa, El Dorado, Fresno, Glenn, Kern, Kings, Madera, Merced, Placer, Sacramento, San Joaquin, Shasta, Stanislaus, Sutter, Tehama, Tulare, Yolo, and Yuba. "San Francisco Bay Area" includes Alameda, Contra Costa, Marin, Napa, San Francisco, San Mateo, Santa Clara, Solano, and Sonoma counties. "Los Angeles" refers to Los Angeles County; "Inland Empire" refers to Riverside and San Bernardino counties; and "Orange/San Diego" refers to Orange and San Diego counties. Residents of other geographic areas are included in the results reported for all adults, but sample sizes for these less populous areas are not large enough to report separately. We also report the results for those who identify as Democrat, Republican, independent, and other voters, but sample sizes for other voters are not large enough to report separately. Lastly, and important to this survey with its focus on jobs and the economy, we reported the results for members of a union and non-union members, as well as adults who have or have not attended a community college.

The overall margin of error is +/- 3.4 percent. The margin of error is calculated at the 95 percent confidence interval. When applicable, we compare the AI survey findings to the 2023 and 2024 Carnegie California Global Affairs Survey results that were conducted with the same methodology and a number of national and regional surveys. This survey references surveys from TechEquity (2025),[1] Ipsos (2024),[2] Politico (2023),[3] Stanford Deliberative Democracy Lab (2024),[4] and MITRE-Harris (2023).[5]

Economy, Work, and the Labor Market

The rapid adoption[6] of AI in the workplace between 2023 and 2024 happened faster than expected, with firms witnessing an annualized growth rate increase of about 78 percent, and individuals reporting an almost 145 percent annualized growth rate.[7] In July 2025, the release of the White House AI Action Plan called for the removal of federal regulations in an effort to further quicken the pace at which AI development and deployment are taking place, potentially enabling an overhaul of how Americans engage in work, the labor market, and, as a result, the economy.

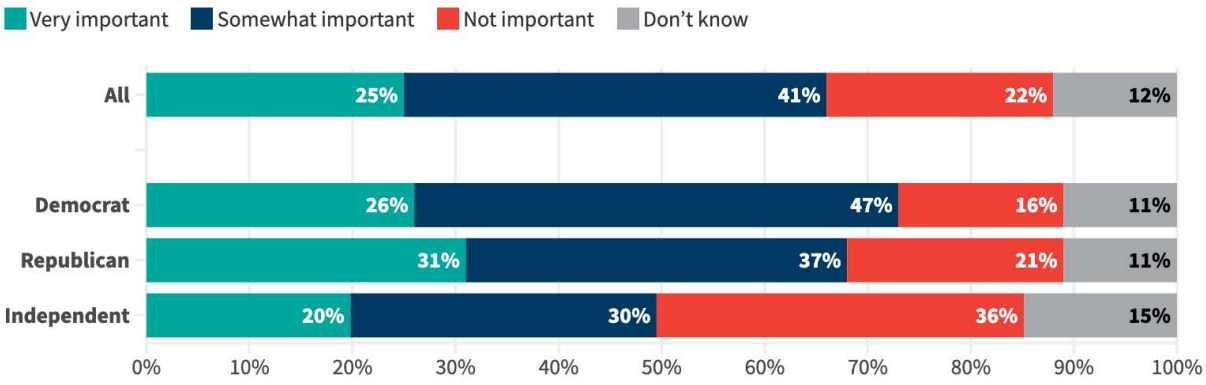
Economic Impacts of AI

Most Californians say that AI is important (27 percent very, 43 percent somewhat) to the nation's economic growth and competitiveness, and a similar share believes that AI is important (25 percent very, 41 percent somewhat) to the state's economic growth and competitiveness. One in five believe that AI is not important to the nation's (20 percent) and the state's (22 percent) economic growth and competitiveness. The perspective that AI is very important for the nation's economic growth and competitiveness is more often held by immigrants (40 percent), full-time workers (36 percent), college graduates (35 percent), those with incomes of over $100,000 (35 percent), and San Francisco Bay Area residents (35 percent). Majorities hold the view that AI is important across age, gender, income, education, racial/ethnic groups, and state regions. These demographic and regional trends are similar for the importance of AI for the state's economic growth and competitiveness.

Notably, majorities across partisan groups think that AI is important for the nation's economic growth and competitiveness (76 percent Democrats, 75 percent Republicans, 56 percent independents). There are similar responses from Democrats and Republicans when asked to rate the importance of AI for the state's economic growth and competitiveness (see Figure 1).

Figure 1. Artificial Intelligence and California Economic Growth

How important do you think AI is to California's economic growth and competitiveness?



Source: 2025 Carnegie California AI Survey
N: 1,601
Note: Individual percentages will not necessarily add up to 100 given rounding.

Californians have mixed views when asked if AI will make California's economy better (27 percent) or worse (31 percent) in the next three years (20 percent same, 22 percent don't know). Fewer than half across partisan and demographic groups and state regions expect that AI will make the California economy better in the next three years. Meanwhile, half of Californians believe that AI will make the job market worse (50 percent) and the gap between the rich and the poor worse (52 percent) in their part of California in the next three years. Pluralities across partisan groups (53 percent Democrats, 43 percent Republicans, and 49 percent independents) and demographic groups and state regions hold the view that AI will make the job market worse in their part of California in the next three years. Partisans vary in their view that AI will make the gap between the rich and the poor worse in their part of California (60 percent Democrats, 36 percent Republicans, 60 percent independents). Recent national surveys have similarly shown the American public to hold a pessimistic view of how AI will shape work in the future.[8]

About two in three adults believe that large companies will benefit from AI (68 percent) compared to about one in three who think that small companies will benefit from AI (35 percent). This result echoes findings from a 2024 survey that found 69 percent of Americans felt that big businesses would benefit from AI.[9] Majorities across all partisan and demographic groups and state regions think that large companies will benefit from AI. Fewer than half across all partisan, demographic, and state regions believe that small companies will benefit from AI. When asked to choose which industry in California will be most impacted by AI in the next three years, the top mentions were engineering, coding, and information technology (34 percent), followed by education and entertainment (10 percent each), and finance, real estate, and healthcare (6 percent each).

Californians have mixed views when asked if AI will make California's economy better or worse in the next three years.

Current Work and AI

AI is not a wholly new technology, nor is its use in the workplace always visible. Nonetheless, Californians have a sense of its extensive deployment in the workplace, while not necessarily feeling they are being prepared for its future uses. Among early career professionals (ages 22–25) in particular, a working paper by the Stanford Digital Economy Lab provides "early, large-scale evidence consistent with the hypothesis that the AI revolution is beginning to have a significant and disproportionate impact on entry-level workers in the American labor market."[10]

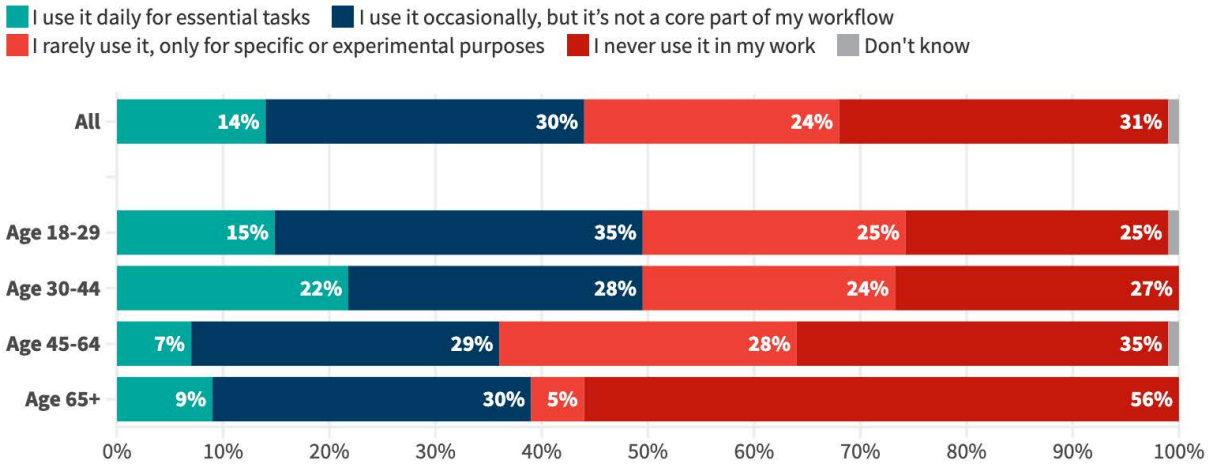
Most employed residents believe that AI is being used in their workplace or industry (11 percent extensively, 34 percent moderately, and 28 percent minimally). Majorities say that AI is being used "extensively" or "moderately" among those who are under 30 years old (56 percent), those who are earning over $100,000 (53 percent), and college graduates (51 percent). While 44 percent say that they use AI at work daily or occasionally, half or more use AI with this level of frequency among those living in the San Francisco Bay Area (55 percent), college graduates (51 percent), and those with incomes over $100,000 (55 percent). Nationally, only 16 percent of workers reported using at least some AI at work in late 2024.[11] A plurality of workers report that AI has led to a "minor" improvement in their productivity at their job (46 percent). Across age groups, employed adults who are under 30 years old are the most likely to say that AI has led to "great" improvement of their work or productivity at their job (36 percent).

Only 5 percent of employed residents say that to date, AI has shifted their career path, and only 10 percent say it has shifted their career focus while they are still in the same field. Fifty-seven percent report that AI has not changed their career plans, although fewer than half say this among workers who are under 45 years old and living in the San Francisco Bay Area.

The majority of employed residents say they are interested (25 percent very, 33 percent somewhat) in being offered training and courses on AI uses in the future at work. Employed residents who are younger (29 percent among those aged 18 to 29; 31 percent among those aged 30 to 44) and live in the San Francisco Bay Area (34 percent) are the most likely of their demographic groups to say they are "very interested" in training and courses on AI uses at work in the future. Despite this interest, only 17 percent have taken classes or received any training on AI tools in the past twelve months. Those most likely to say that they have had classes or been offered training are under 30 years old, college graduates, those earning over $100,000, and big city residents. National trends track the limited access to training to date. A poll earlier this year found that only 28 percent of employed U.S. adults had been offered training in AI use at their jobs.[12]

Figure 2. Current Use of Artificial Intelligence at Work

How often do you use AI for work?



Source: 2025 Carnegie California AI Survey
N= 557
Note: Individual percentages will not necessarily add up to 100 given rounding.

Future Work and AI

California is developing programs to support a new generation of workers. In early August 2025, Governor Gavin Newsom announced an agreement between the state government and four of the largest AI companies to equip community colleges and California State University systems with tools to train a new AI workforce.[13] Many Californians see AI playing a growing role in future work, and while most are not concerned about its short-term impacts on their employment, their views are more pessimistic about their long-term job prospects due to AI.

About half of California's employed adults (48 percent) expect to be using AI more in the future at work. Majorities of those under 45 years of age, men, college graduates, big city residents, and both Democrats and Republicans hold this view. Across racial and ethnic groups, Hispanic respondents (36 percent) are the least likely to expect to be using AI more in the future, while majorities of Asian, Black, and White respondents say they expect to be using AI more.

The regional differences here are notable. San Francisco Bay Area residents (61 percent) are the most likely to say that they expect to be using AI more at work in the future, while fewer than half of those living in the Inland Empire, Central Valley, Los Angeles, and Orange/San Diego counties hold this view (see Figure 3).

About three in four employed residents think that skills to understand and use AI will be important in today's economy. However, about half believe that AI in the workplace will lead to fewer job opportunities in the long run.

Three in ten employed adults are concerned (10 percent very, 21 percent somewhat) about losing their job due to being replaced by AI in the next three years, while six in ten are not too concerned (32 percent) or not at all concerned (32 percent) about this possibility. A 2024 Public Policy Institute of California survey showed similar findings (9 percent very, 24 percent somewhat).[14] Relatively few say they are very concerned about losing their job across partisan and demographic groups and state regions.

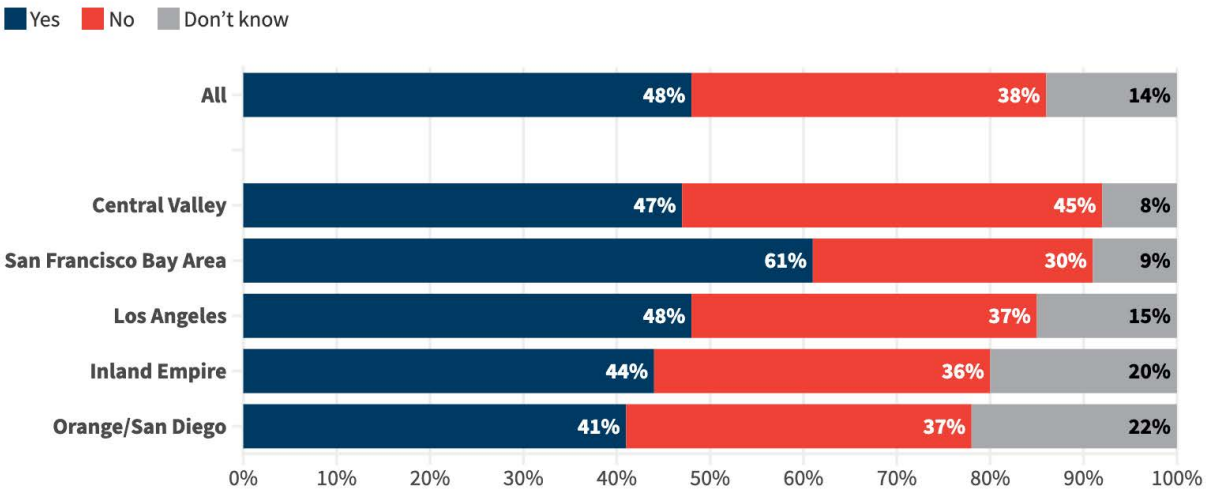
About three in four employed residents think that skills to understand and use AI will be important for a worker to be successful in today's economy. Only one in ten say AI skills will not be important or not important at all for future success. Across partisan and demographic groups and state regions, most believe that skills to understand and use AI will be important.

However, about half believe that the use of AI in the workplace will lead to fewer job opportunities for themselves in the long run. Only 8 percent believe that the use of AI will lead to more job opportunities for themselves in the long run. Pluralities hold a pessimistic view about future job opportunities across partisan and demographic groups and state regions. Those who have lower incomes and rent their houses are more likely to believe that the use of AI will reduce their job opportunities, as opposed to those with higher incomes (over $50,000) and who own their house. Less than a year ago, a national survey had similar findings in that only 6 percent of employed adults believed that the use of AI will lead to more job opportunities.[15] In 2023, Hollywood writers from the Writers Guild of America and actors from the Screen Actors Guild–American Federation of Television and Radio Artists went on strike in an effort to, among other issues, protect their jobs from AI.[16] Among the regions surveyed, Los Angeles had the highest percentage (57 percent) who believe they would have fewer job opportunities in the future as a result of AI.

Statewide, 69 percent of Californians say they would support public policy protecting workers from job displacement caused by AI, with 50 percent strongly supportive, 19 percent somewhat supportive, and a further 11 percent whose views depend on the type of protection. Notably, given the research positing a disproportionate impact on entry-level jobs, this support is most intense among younger Californians. Fifty-eight percent of 18–29-year-olds are strongly supportive, compared with around half of respondents age 30 and older.[17] Support also varies by party affiliation: 79 percent of Democrats support protections (60 percent strongly), compared with 58 percent of Republicans (35 percent strongly) and 61 percent of independents (49 percent strongly).

Figure 3. Future of Artificial Intelligence Use at Work

Do you personally expect to be using AI tools more in the future at work?



Source: 2025 Carnegie California AI Survey
N= 758
Note: Individual percentages will not necessarily add up to 100 given rounding.

Privacy, Surveillance, Bias, and Harm

Since passing landmark privacy legislation in 2018,[18] California has been at the forefront of U.S. efforts to regulate businesses' collection and use of personal information. Likewise, the state has enacted numerous policies aimed at AI risk reduction. Nationally, there are thriving debates[19] over the lack of federal privacy standards and the appropriate balance of state and federal responsibility for safeguarding Americans from privacy intrusion and other harms in the age of AI.

As the epicenter of global AI development and home to a disproportionate share of technologists and executives shaping AI systems, California has a unique standing, and Californians have a distinct vantage point on the AI revolution. As the world's fourth-largest economy and a globally consequential regulator in its own right, California's policymakers have a unique capacity to impact the development and spread of AI systems. Californians' views on fundamental topics like privacy, fairness, and the potential drawbacks of AI systems therefore have resonance not only in Sacramento but far beyond the state's expansive borders.

On Privacy

For Californians, privacy remains a vital issue. Seventy-two percent of Californians rank privacy intrusion among the AI-related risks that cause them the greatest concern, making it the most broadly worried-about category of risk surveyed. Californians express significant

A solid majority of respondents do not "trust that companies will build and use [AI] systems in ways that protect [their] personal data."

misgivings about their knowledge of and control over AI companies' use of personal information. One-third "strongly" or "somewhat" agree that they "know what personal information about [themselves] is used or shared by AI systems," while two-thirds "somewhat" or "strongly" disagree. Even fewer Californians—some 20 percent—believe they have agency over how their information is used. By contrast, 65 percent report disagreeing that they "can control what personal information about [themselves] is used or shared by AI systems."

Nor are Californians comfortable that AI companies can be trusted to protect user privacy. A solid majority of respondents (59 percent) do not "trust that companies will build and use [AI] systems in ways that protect [their] personal data," with a plurality (37 percent) reporting that they "strongly disagree" that companies can be trusted. As we will see, the survey results highlight at least one prescription responsive to these worries.

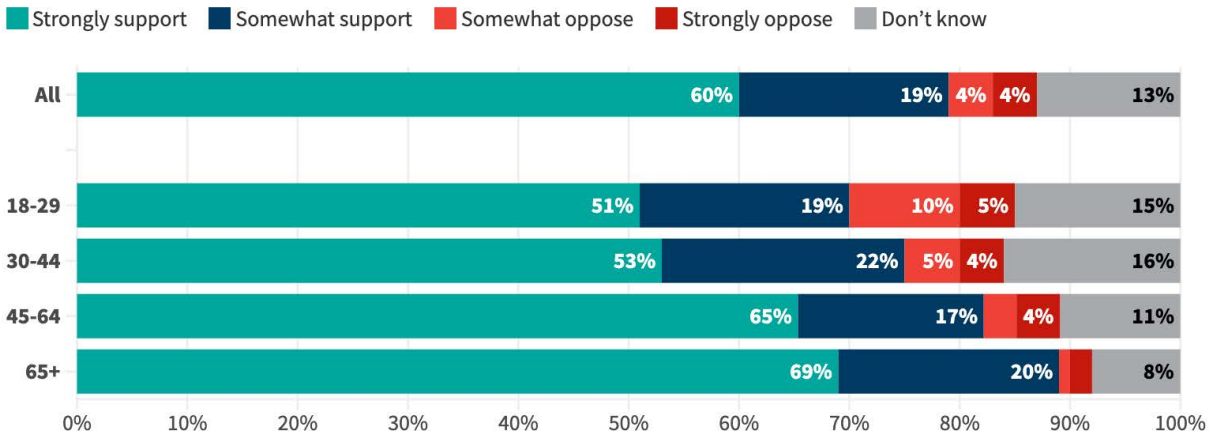
Transparency in Support of Privacy and Legal Compliance

Fueled by worries about their privacy, agency, and trust for AI companies, Californians broadly support transparency, both in company disclosures about the processing of personal information and in legal protection for whistleblowers to speak up regarding potential violations of law. Sixty-five percent "strongly agree" that companies should be required to tell people the types and sources of their personal information that an automated decision-making system analyzes when making important decisions about them, such as determining their access to lending, insurance, housing, education, or employment opportunities. A further 18 percent (totaling 83 percent collectively) say they "somewhat agree." Similar margins believe (80 percent total, with 59 percent "strongly" agreeing) that companies should be required to tell people which of their interests, preferences, behaviors, or other personal traits are analyzed or used by automated decisionmaking systems to make important choices about them.

Californians' preference for transparency is not limited to requiring companies to provide information. It extends as well to protecting AI company insiders who choose to speak up about potential violations of the law. Seventy-nine percent of Californians support having a state policy to protect whistleblowers at technology companies who speak out about their employers using AI in ways that violate the law, with 60 percent "strongly" supportive and only 8 percent "strongly" or "somewhat" opposed. Democrats and Republicans report slightly differing intensities of support, with 71 percent of Democrats and 56 percent of Republicans "strongly" in favor. Still, the overall proportion of supporters shows striking commonality across partisan lines (86 percent of Democrats and 84 percent of Republicans). Support is markedly stronger among registered voters than unregistered, with 83 versus 60

Figure 4. Protecting Whistleblowers

Do you support or oppose having a state policy to protect whistleblowers at the technology companies when they speak out about their employers using artificial intelligence in ways that violate the law?



Source: 2025 Carnegie California AI Survey
N= 557
Note: Individual percentages will not necessarily add up to 100 given rounding.

percent at least somewhat supportive. There are also notable differences across age groups (see Figure 4). These findings bear particular relevance at a time when California has just adopted legislation extending whistleblower protection to employees who raise concerns about catastrophic risks or violations of law involving foundation models.

Notwithstanding their support of transparency, Californians also have views on the ways in which their information is used: by whom, for what purpose, and in what accord with their own expectations and permission.

AI-Enabled Uses of Personal Information

As a general matter, Californians express discomfort with how their personal information is used. However, limited gradations of trust are visible depending on whose systems are at work. Only a quarter of Californians are "very" or "somewhat" comfortable with privacy technology companies using their personal information. Sixty-three percent are uncomfortable. They are even more uncomfortable with the government using their personal information: 76 percent report they are "not comfortable," with only 18 percent "very" or "somewhat" comfortable. This discomfort is slightly lower when a government partners with private companies; but even then, very large majorities (72 percent) still report being "not comfortable," compared with 20 percent expressing comfort and a tiny fraction (3 percent) saying they are "very comfortable." While regional variation is limited, San Francisco Bay Area residents are slightly more comfortable with private companies' use of their personal

More than two-thirds of Californians are concerned with AI systems accessing content they have shared, even publicly for unrestricted distribution, such as posts, messages, photos, or videos.

data (33 percent versus 25 percent statewide are at least "somewhat" comfortable). This does not extend to government usage, however: Bay Area residents are less comfortable than the statewide average, albeit within the margin of error.

In at least some environments, this discomfort is allayed by disclosure and consent. When asked about AI accessing users' activity to personalize interactions, a majority of Californians say that "AI chatbots should prioritize responses that do not rely on user data." While only 7 percent seem broadly comfortable with default personalization, a further 41 percent believe that chatbots should use additional data sources, such as the user's online activity, to help personalize their interactions, provided they do so with the user's permission.

Notably, given the training of generative AI models on a wide range of digital content,[20] Californians expressed concern with AI systems accessing content they have shared, even publicly for unrestricted distribution, such as posts, messages, photos, or videos. Amid continuing controversy over AI developers' use of sources like social media posts,[21] or videos,[22] only 6 percent of Californians are "very comfortable" with AI systems accessing content they have posted publicly, while more than two-thirds (69 percent) say they are "not comfortable" with such access.

Unsurprisingly, an even larger majority of Californians (79 percent) express discomfort with the idea of AI systems accessing content they have shared *privately*, such as information shared directly with friends or to a limited, access-restricted audience. Interestingly, the fraction of Californians who are "very comfortable" with AI systems accessing their posts, messages, photos, or videos is the same (6 percent) regardless of whether the content has been made publicly available or kept private by the user. While highly unrepresentative of Californians' overall concern for AI systems' use of user-generated content, this suggests there is a cohort of Californians, albeit a very small one, with a highly permissive view of the issue.

An even broader majority of Californians (81 percent) are uncomfortable with the idea of AI systems *sharing* information about them, such as posts, messages, photos, or videos, with other users. AI companies have warned that user conversations can be stored and subject to compelled disclosure, for example, in legal proceedings, and observers have noted the potential for AI models to regurgitate proprietary or personal information,[23] even if inadvertently.[24] These survey findings suggest a gap between Californians' expectations and current reality, though it remains to be seen whether consumer expectations will drive changes in policy or model behavior or, alternatively, will adjust over time as users acclimate to new privacy norms shaped by the realities of model and regulatory design.

Facial Recognition and Bias

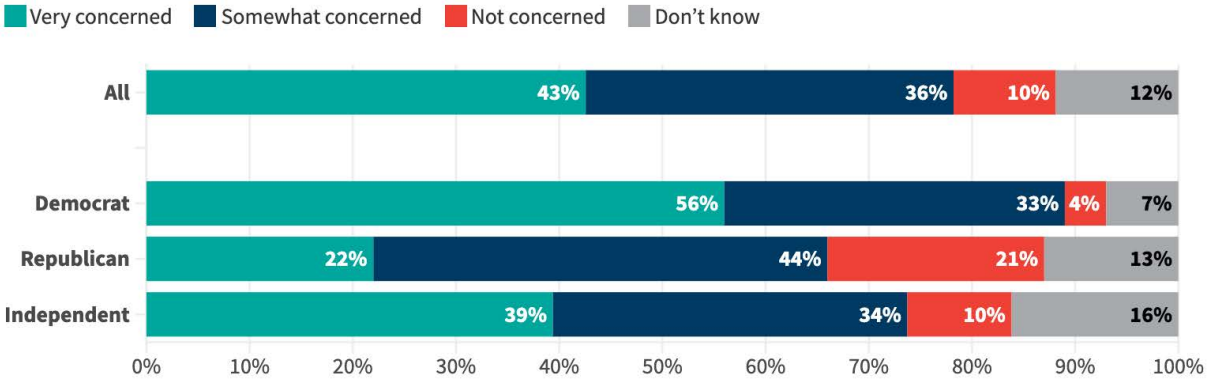
Interestingly, Californians expressed more mixed views on the concrete but charged issue of facial recognition. A majority of respondents are open to at least some deployment by private companies. Barely a quarter (27 percent) were "very comfortable" or "somewhat comfortable" with the idea of businesses using facial recognition tools in the abstract, but a further 28 percent say "it depends on the purpose (e.g., security vs. marketing)," possibly reflecting the truism that people are often more comfortable having their data processed for a tangible benefit.

Californians express somewhat greater comfort with law enforcement deploying facial recognition. A plurality (44 percent combined) are either "very comfortable" or "somewhat comfortable" with such use "to identify suspected offenders," while another 22 percent are open to it, depending on a more specific assessment of purpose. This suggests that, notwithstanding risks such as misidentification or differential impact across communities, 66 percent of Californians have some openness to facial recognition's use in the public safety context.

Still, Californians express significant concern with the potential for AI systems to behave in biased or discriminatory ways (see Figure 5). Seventy-nine percent of Californians express concern that AI systems might "disfavor people from disadvantaged backgrounds or reinforce existing biases." Of these, a plurality (43 percent) are "very concerned," and a further

Figure 5. Concern over Biased Outcomes

How concerned are you that AI systems, trained on historical data and used for important decisions (like hiring, loan requests, rental applications), might disfavor people from disadvantaged backgrounds or reinforce existing biases?



Source: 2025 Carnegie California AI Survey
N= 1,601
Note: Individual percentages will not necessarily add up to 100 given rounding.

12 percent "don't know," leaving only 10 percent who say they are "not concerned" with the prospect. Perhaps unsurprisingly, given the polarized national debate over issues of diversity and equity, these findings exhibit significant political differentiation. Eighty-nine percent of Democrats are at least somewhat concerned, compared with 66 percent of Republicans: still a strong majority, albeit a less ardent one (44 percent of Republicans are "somewhat" and 22 percent "very" concerned). Independents fall somewhat in between, with 34 percent being "somewhat" and 39 percent "very concerned."

Bias is one frequently cited risk of algorithmic decisionmaking and AI model deployment. The survey sheds light on Californians' views about a number of other risks as well.

On Risks Posed by AI Systems

Turning to a broader taxonomy of AI's potential risks, as we have seen, privacy was the most frequently cited category of concern (72 percent). While strong majorities worry about privacy across every age bracket sampled, these worries are broadest among older respondents, from slightly less than two-thirds among 18–29 and 30–44-year-olds to 76 percent among 45–64-year-olds and 82 percent among respondents aged 65 years and above. Across every risk category sampled (apart from "don't know"), Californians age 65 and above report the highest percentage of concern, though the magnitude of this effect varies by risk. For example, while there is only a 3 percent difference between 18–29-year-olds and those 65 and older in their worry about AI's climate impacts, Californians age 65 and older report serious worry about AI's impact on elections at a rate 32 percent greater than those age 18–29.

Across age groups, however, strong majorities (greater than 60 percent) worry about a number of risks beyond privacy. These include misuse of AI for spam or fraud (65 percent), misinformation or use to manipulate others (64 percent), and cybersecurity harms (61 percent).

Majorities also ranked among their top concerns the spread of false content, such as audio or video deepfakes (58 percent); loss of human control of AI (56 percent); use of AI to influence elections (55 percent); and use of AI to manufacture or spread harmful content, such as nonconsensual intimate imagery, harassing or bullying content, hate speech, or the promotion or incitement of violence (53 percent). Slightly less than a majority of respondents cited concerns over job loss due to automation (49 percent); algorithmic bias or discrimination (48 percent); generation or spread of unvetted or inaccurate medical information (48 percent); and hallucination, in which AI models generate factually unfounded responses (47 percent).

Despite significant debate within the AI community over proposed California legislation seeking to curb catastrophic risk,[25] such as the AI-assisted proliferation of weapons of mass destruction, a smaller proportion of respondents (38 percent) cited AI's use to help build, acquire, or use chemical, biological, or radiological weapons among their biggest worries. Likewise, climate and environmental impacts and rising energy costs to power data centers

were each cited by just over a third (35 percent) of respondents. Whether these relatively lower prioritizations reflect Californians' considered judgment or other factors, such as lower awareness of these issues, would require more information to determine.

Government and Democracy

AI's impact on government and democracy is nascent, raising both risks and opportunities. Federal, state, and local agencies are working to mitigate risks in a landscape of evolving regulation while also experimenting with emerging technologies to improve public services and government efficiency. A national 2024 Ernst & Young Survey found that 64 percent of U.S. federal government employees and 51 percent of state and local government employees reported "using an AI application daily or several times a week," with reported use in areas including border patrol, drone manufacturing, biometric data collection, and more.[26] As AI use is on the rise, many report a desire for requisite training, with calls to upskill workplaces on AI coinciding with demands for regulatory guardrails to protect citizen data and mitigate bias.

While California has been at the forefront of various efforts to regulate and use AI, its residents have mixed views about its influence on government, Many report hesitancy and limited knowledge about AI's role in politics and government, underscoring its emerging influence even in a technology-leading state. Across demographics, regions, and partisan lines, respondents generally prize safety over innovation and are hesitant about AI adoption by government and its impact on elections. At the same time, substantial numbers of Californians report using digital tools to access public services, especially health services where about a quarter of Californians report using digital tools, and about a quarter of Californians also report optimism about AI's potential to improve policy and democratic processes.

AI and Accessing Services

Few Californians say AI has improved their interactions with government, highlighting gaps in its use and/or communication about its potential use across the state. Only a small number (4 percent) reported that AI "significantly improved" their ability to access public/government services, while a larger number (13 percent) reported their access had "slightly improved." The most common answer among respondents (37 percent) was that they "don't know."

At the same time, Californians are using technology in notable ways to interact with government services, indicating potential areas for further application of AI in the future, most significantly for health and wellness (see Figure 6). A substantial minority (23 percent)

Across use cases, more men reported using digital tools to access services than women.

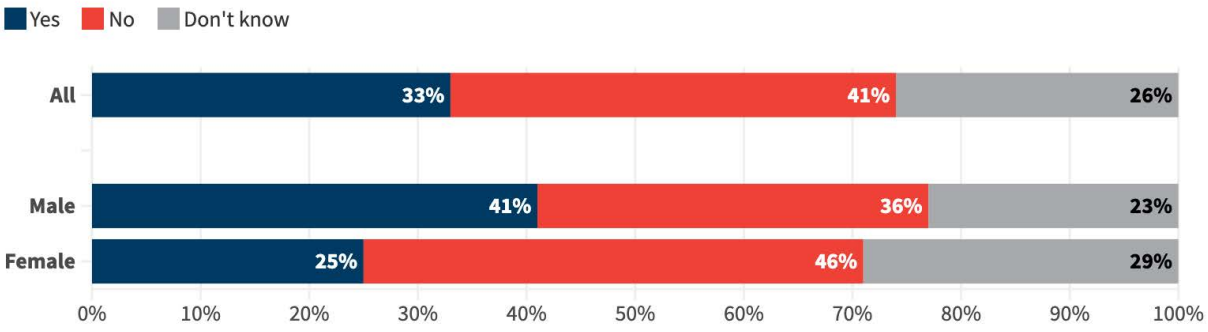
report having used digital tools to access information about health services, while fewer report having used AI for information about transportation (14 percent) and elections (13 percent), and even fewer for legal information (10 percent).

Californians' use of digital tools for public services differs based on age and gender, indicating a need to identify and address gaps in trust and access. Across use cases, more men reported using digital tools to access services than women. More men (26 percent) reported using digital tools to access information about health services than women (21 percent). Similarly, though Californians' use of AI for information about voting and elections is lower than for information about health services, more men (16 percent) reported using digital tools to access election information than women (10 percent).

Younger Californians tend to use digital tools more than older generations, though people of different ages use digital health services more evenly than other services. Californians aged 18–29 report only slightly higher use of digital tools for health services than those aged 45–65 (28 percent to 22 percent, respectively). However, for accessing election information, the difference is more significant (21 percent to 9 percent).

Figure 6. Views on AI and Civic Engagement

Do you think that AI tools can help you to become a more informed and engaged voter and citizen?



Source: 2025 Carnegie California AI Survey
N= 758
Note: Individual percentages will not necessarily add up to 100 given rounding.

Government Use of AI

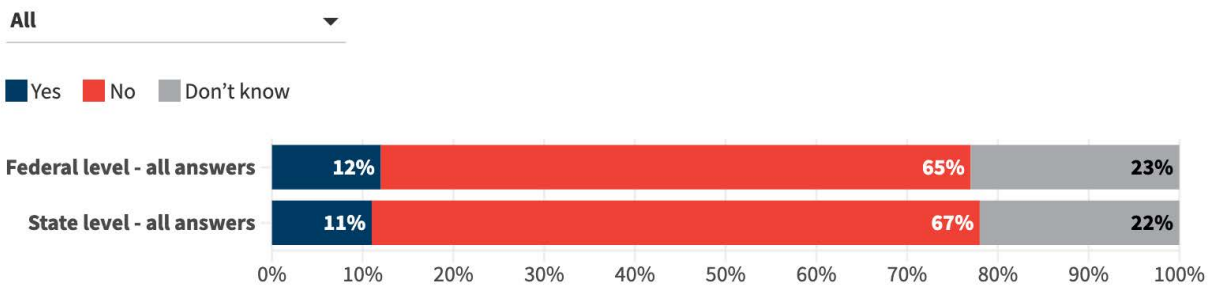
As noted above, Californians are generally skeptical about government use of AI (see Figure 7). A majority (65 percent) believe that federal government agencies should not use AI to make decisions that directly affect them and their community, though a minority agree with federal government AI use (12 percent) and 23 percent said they don't know if they agree or not, reflecting uncertainties about AI technology use by government and its potential influence on communities.

Unease about government use of AI spans the political spectrum. A slightly higher percentage of California's Democrats (67 percent) think that the federal government should not use AI in decisions that affect them than Republicans (62 percent). In fact, gender divides are more notable than partisan divides. More men believe the federal government should use AI for decisions affecting them and their communities (16 percent) than women (8 percent).

Californians report similar skepticism about government use of AI at the state and local levels. Most do not want state and local governments to use AI for decisions directly affecting them and their communities (67 percent), while a small minority do (11 percent). But again, notable numbers (22 percent) say that they simply "don't know." Respondents have slightly more confidence in the use of AI by local police, fire departments, and emergency services (19 percent support its use). A gender divide is again visible over this application of AI for safety and emergency services—more men support AI use for this purpose (23 percent) than women (15 percent).

Figure 7. Federal and State Government and AI

Do you believe that government agencies should use AI to make decisions that directly affect you and your community?



Source: 2025 Carnegie California AI Survey
N=1,601
Note: Individual percentages will not necessarily add up to 100 given rounding.

When asked about AI's use to improve the efficiency of government, however, Californians are more supportive. A large share of Californians say AI should be used to improve the efficiency of government work, with more saying it is "somewhat important" (36 percent) than those who say it is "not important" (28 percent). Still about one in four report that they "don't know" (23 percent), again signaling uncertainties among the public about AI technologies and their effects.

Mixed views emerge when asked about AI's potential use by the government to address crime. Californians are split on the use of AI for this purpose. The same number are optimistic about AI's use to address crime and advance justice (35 percent) as are pessimistic (35 percent). The rest say they don't know (30 percent).

Although some Californians report optimism about AI's potential to help reduce crime, most indicate concern about guardrails around how data is used. Most Californians report some level of concern about AI use for government surveillance and data collection, with many reporting being "very concerned" (54 percent) and still others report being "somewhat concerned" (27 percent).

AI's Influence on Democracy and Civic Life

The influence of AI-generated content on democracy is an important concern for Californians. Fifty percent of Californians say that they are "not confident" they can tell the difference between real and AI-generated information. Only a small number report that they are "very confident" that they can tell the difference (8 percent), and some report being "somewhat confident" (32 percent).

The potential for such content to influence elections concerns respondents. Fifty-seven percent of Californians report that they are "very concerned" about the influence of deepfakes and other AI-generated content in elections. Similar numbers report being "very concerned" about AI-generated content online heightening political violence and polarization (55 percent).

Fifty percent of Californians say that they are "not confident" they can tell the difference between real and AI-generated information.

Californians are split, however, when it comes to how AI can help democracy and governance. Nearly a third of Californians think that emerging technologies will help drive productive policy outcomes (32 percent), while similar numbers report that they do not think technologies like AI will help (33 percent). A similar, but slightly higher, number reports that they don't know (36 percent). Responses are nearly identical to this question among Democrats and Republicans, revealing similarly unsettled views among Californians around AI's influence on democracy, regardless of party.

Similarly, split views are visible among Californians about AI's potential to play a productive role in democratic processes, although here Californians are slightly more pessimistic. Twenty-three percent of Californians think that emerging technologies can help in this regard, while higher numbers think emerging technologies will not play a productive role (47 percent), and substantial numbers report that they don't know (31 percent). On voting, responses are slightly more optimistic. Thirty-three percent of Californians think AI can enable them to become more informed and engaged voters and citizens, while more say it will not help (41 percent), and a smaller number say they don't know (26 percent). Responses are similar across party affiliation, but some variance is visible based on geography (more San Francisco Bay Area and LA residents report optimism that AI can help them as voters and citizens—37 percent and 35 percent, respectively—than those in the Central Valley—28 percent).

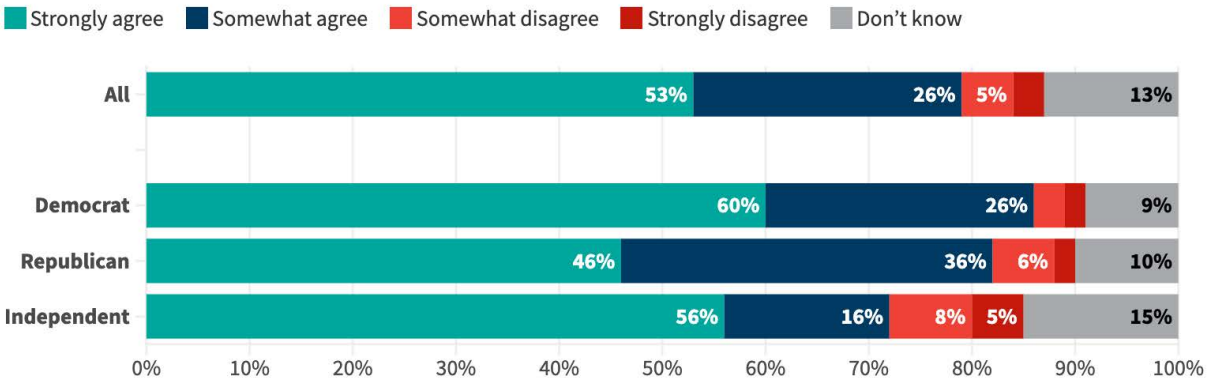
But this question reveals a major gender divide: 41 percent of men think AI will support them as informed voters, compared with only 25 percent of women.

AI Regulation and Education

Most Californians report concern about safety when it comes to AI, indicating a potential basis for more robust regulation over its development and use. Fifty-three percent strongly agree that safety should be prioritized over innovation, while only 3 percent strongly disagree (see Figure 8). The split is similar across men and women and relatively similar across regions and age groups, although there is a slight increase in interest in safety over innovation emerges among the oldest group (63 percent strongly support safety over innovation among those aged 65 and up as compared with around 50 percent for all other age groups).

Figure 8. Support for Safety Over Innovation

Do you agree or disagree with this statement: "On balance, safety should be prioritized over innovation, when it comes to AI regulation?"



Source: 2025 Carnegie California AI Survey
N=1,599
Note: Individual percentages will not necessarily add up to 100 given rounding.

Here there is a notable partisan divide. Californians' views on the balance of safety and innovation differ somewhat based on political party. More Democrats agree that safety should be prioritized over innovation (60 percent) than Republicans (46 percent).

Opinions also differ on the question of safety and innovation when considering race. More White respondents strongly agree that safety should be prioritized over innovation (62 percent) than Black respondents (43 percent) and Hispanic respondents (47 percent).

With the fast deployment of AI across society, Californians generally support recent measures to mandate AI education in schools. More Californians (46 percent) support a new California policy to advance AI literacy in schools than oppose it (25 percent). Within this, more men (53 percent) support mandated AI education than women (38 percent). Support for the policy is slightly higher among Democrats (53 percent) than Republicans (48 percent). Support for AI education in schools is highest in the San Francisco Bay Area (56 percent) compared with other areas (43 percent reported support in all other regions, inclusive of Central Valley, Los Angeles, Inland Empire, and Orange/San Diego). Responses show that nearly half the state favors investing in an effort to help future Californians navigate an economy and government increasingly influenced by AI.

Safety and Public Options

The potential impacts of AI across society include catastrophic risks, including risks to critical infrastructure, increased cyber offensive capabilities, and the proliferation of biological weapons.[27] These potential capabilities, while not yet emergent, have been the focus of diplomats, scientists, and national policymakers.[28] The inaugural convening of the International Network of AI Safety Institutes in San Francisco in November 2024 included AI safety experts from eleven countries. The 2025 International AI Safety Report, authored by more than 100 scientists in a fashion similar to the Intergovernmental Panel on Climate Change, regarded such risks that "some experts think that such risks are decades away, while others think that general-purpose AI could lead to societal-scale harm within the next few years."[29] Released in June, the California Report on Frontier AI Policy,[30] featuring input from scholars from Carnegie, Stanford University, and UC Berkeley, noted the importance of early design and policy choices in emerging technologies. It also noted that "greater transparency, given current information deficits, can advance accountability, competition, and public trust as part of a trust-but-verify approach."

A bipartisan majority of Californians agree that making AI safe and secure for public use needs to be a nationwide effort across civil society, industry, government, and universities.

A majority (67 percent) of Californians agree that making AI safe and secure for public use needs to be a nationwide effort across civil society, industry, government, and universities. Results indicate bipartisan support, as Democrats and Republicans are split by a small margin, with 76 percent of Democrats (51 percent strongly, 25 percent somewhat) and 68 percent of Republicans (36 percent strongly, 32 percent somewhat) in agreement with the statement. In contrast, older Californians are more likely to believe in the multistakeholder effort; 97 percent of Californians ages 45 and older are in strong agreement, whereas only 69 percent of Californians younger than 45 feel the same.

In terms of the involvement of industry, the majority (72 percent) of Californians agree that organizations investing in AI tools should also help establish guidelines for AI's safe and responsible use. As for the government, 77 percent (57 percent strongly, 20 percent somewhat) of Californians agree that the government should require AI companies to test their most advanced systems for safety and provide a detailed plan for how they'll prevent harm.

Public Option

The high cost of developing AI, and especially generative AI such as large language models, combined with their market-oriented nature, has led many governments to consider pursuing public options or computing reserved for smaller businesses and researchers. In California, the vetoed SB-1047 and SB-53 both mandated the creation of public AI known as CalCompute. Other subnational jurisdictions, including New York (Empire AI Consortium), as well as nations like the United Kingdom and India, have explored or even begun to build out such efforts. Just as the frontier models remain relatively new, so too do such public AI efforts, raising as they do knotty questions around funding and access.

A plurality (45 percent) of Californians support the state of California creating a publicly available, shared virtual platform (public cloud computing cluster) to advance artificial intelligence research and development, with the goal of ensuring AI is safe, ethical, equitable, and sustainable for all. Partisan lines are divided on this initiative, with a majority of Democrats (56 percent) in support and a plurality of Republicans (42 percent) in opposition. Teachers and other education professionals (64 percent), scientists (61 percent), and healthcare providers (56 percent) were among the top choices for whom Californians believe would get the most out of a publicly available shared online platform.

Looking Forward

An early 2022 paper by the AI company Anthropic noted that LLMs are both highly predictable and highly unpredictable.[31] The history of the technology's development since then has borne this out. AI comes with surprises.[32] Sound policymaking, the California Report on Frontier AI Safety noted,[33] must be adaptable to such surprises.[34]

As Carnegie California's first-ever statewide AI survey shows, Californians believe that AI will significantly impact their work, communities, and democracy, but that general point is balanced by high levels of anxiety and uncertainty around specific impacts. Surprises will occur, but the demand signals for enhanced transparency, training, and education around AI, combined with Californians' measured optimism around evidence-based policymaking, are already clear.

Advisers

Stephen Caines Esq., chief innovation officer and budget director, City of San José

Lilian Coral, vice president of technology and democracy programs, New America

Elena Cryst, director of policy and society, Stanford Institute for Human-Centered Artificial Intelligence

Graham Drake, senior policy advisor, Tony Blair Institute for Global Change

Mitra Ebadolahi, senior project director, Upturn

Samantha Gordon, chief advocacy officer, TechEquity

David Graham, policy designer in residence, The Design Lab at UC San Diego

Kiran Jain, chief legal officer and board secretary, Replica

Alexander Kapur, founder and CEO, PurposeBuilt

Meredith M. Lee, head of strategic partnerships and chief technical advisor, UC Berkeley College of Computing, Data Science, and Society

Paris McCoy, executive director, TEC LEIMERT

Erikan Obotetukudo, founder, Audacity Assets

Ben Polsky, international strategy fellow, Special Competitive Studies Project (SCSP)

James Regan, former deputy secretary for workforce development, California Government Operations Agency

Robert Rodriguez, managing partner and founder, BarronKent Family Office

Kimberly Rosenberger, former senior government relations advocate, SEIU California

Matthew Scherer, senior policy counsel, workers' rights and technology, Center for Democracy & Technology

Scott Singer, fellow, Technology and International Affairs Program, Carnegie Endowment for International Peace

About the Authors

Ian Klaus is the founding director of Carnegie California, the West Coast office and program of the Carnegie Endowment for International Peace.

Mark Baldassare is a nonresident scholar at Carnegie California. He is survey director at the Public Policy Institute of California, where he holds the Arjay and Frances Fearing Miller Chair in Public Policy. He is also a senior fellow at the Bedrosian Center on Governance in the Sol Price School of Public Policy at the University of Southern California.

Rachel George is a consultant at Carnegie California. She is also a lecturer in international relations at Stanford University, where she teaches about AI, international law, and global affairs. She is a senior fellow for the Research on International Policy Implementation Lab and a fellow at the Georgetown Institute for Women, Peace, and Security.

Scott Kohler is a nonresident scholar at Carnegie California. His research explores the nexus of technology, law, and public policy, with a focus on evolving approaches to regulation and structures of governance. Kohler's recent work has particularly explored legislative debates over AI safety and the broader role of U.S. states in AI policymaking. He has previously worked at Google and Nest Labs and served as a senior legal advisor at the U.S. Department of Energy.

Marissa Jordan is the program manager of Carnegie California at the Carnegie Endowment for International Peace. Since 2018, she has been actively involved in various initiatives at Carnegie, including the Democracy, Conflict, and Governance Program, the Europe Program, the Global Order and Institutions Program, the Sustainability, Climate, and Geopolitics Program, as well as Carnegie's artificial intelligence research.

Abigail Manalese is an intern with Carnegie California at the Carnegie Endowment for International Peace. She is a recent graduate of the University of California, Los Angeles, where her research focused on intersections between international human rights law, migration, and refugee policy. Previously, she served as a research coordinator for student-led ethnographic projects at Foothill College, a research assistant for a project on organizational behavior and policymaking at Stanford University, and as a Global Affairs Fellow with Meridian International Center in D.C.

Acknowledgments

We would like to thank Alexander Marsolais, Kara Eytcheson, and Sarah Higdon of YouGov for conducting the survey with immense care and rigor. Special thank you to Sarah Camacho, research assistant for Carnegie's Asia Program, for her Spanish translation. Thank you to Carnegie's communications team: Alana Brase, managing editor; Helena Jordheim, assistant editor; Amy Mellon, senior graphic designer; and Jocelyn Soly, creative director, for their amazing work on editing this paper and creating the incredible graphics to go along with it. This work is supported by a grant from Omidyar Network.

Notes

1 ”How Californians Feel About AI – Findings From the 2025 AI Compass,” TechEquity, August 19, 2025, https://techequity.us/2025/08/19/how-californians-feel-about-ai/.

2 “Where Americans Stand on AI,” Ipsos, January 31, 2025, https://www.ipsos.com/en-us/where-americans-stand-ai?utm_source=substack&utm_medium=email.

3 “National Tracking Poll, Topline Report,” *Politico*, December 19, 2023, https://www.politico.com/f/?id=0000018c-89a0-d32e-a59d-cbacaf590000.

4 Samuel Chang, Estelle Ciesla, Michael Finch, et al. “Meta Community Forum, Results Analysis,” Stanford Deliberative Democracy Lab, April 2024, https://fsi9-prod.s3.us-west-1.amazonaws.com/s3fs-public/2024-03/meta_ai_final_report_2024-04_v28.pdf.

5 “MITRE-Harris Poll Survey on AI Trends,” MITRE Corporation, September 19, 2023, https://www.mitre.org/sites/default/files/2023-09/PR-23-2865-MITRE-Harris-Poll-Survey-on-AI-Trends.pdf.

6 The meaning of adoption in this case is broad; ranging from any general AI use at work to specific tools within the workplace, as the statistics are drawn from a paper reviewing sixteen different surveys all varying in their definitions of “adoption.”

7 Leland Crane, Michael Green, and Paul Soto, “Measuring AI Uptake in the Workplace,” Federal Reserve, February 5, 2025, https://www.federalreserve.gov/econres/notes/feds-notes/measuring-ai-uptake-in-the-workplace-20240205.html.

8 Colleen McClain, Brian Kennedy, Jeffrey Gottfried, Monica Anderson, and Giancarlo Pasquini, “How the U.S. Public and AI Experts View Artificial Intelligence,” Pew Research Center, April 3, 2025, https://www.pewresearch.org/internet/2025/04/03/how-the-us-public-and-ai-experts-view-artificial-intelligence/.

9 “Where Americans Stand on AI,” Ipsos, January 31, 2025, https://www.ipsos.com/en-us/where-americans-stand-ai?utm_source=substack&utm_medium=email.

10 Erik Brynjolfsson, Bharat Chandar, and Ruyu Chen, “Canaries in the Coal Mine? Six Facts about the Recent Employment Effects of Artificial Intelligence,” Stanford Digital Economy Lab, August 2025, https://digitaleconomy.stanford.edu/wp-content/uploads/2025/08/Canaries_BrynjolfssonChandarChen.pdf?utm_source=twtw&utm_content=20250829&utm_medium=email&utm_campaign=TWTW2025Aug29&utm_term=TWTW%20and%20staff.

11 Luona Lin and Kim Parker, "U.S. Workers Are More Worried Than Hopeful About Future AI Use in the Workplace," Pew Research Center, February 25, 2025, https://www.pewresearch.org/social-trends/2025/02/25/u-s-workers-are-more-worried-than-hopeful-about-future-ai-use-in-the-workplace/.

12 "Americans Still Have A Complicated Relationship with AI," Ipsos, January 31, 2025, https://www.ipsos.com/en-us/americans-still-have-complicated-relationship-ai.

13 "Governor Newsom Partners with World's Leading Tech Companies to Prepare Californians for AI Future," Governor of California, August 7, 2025, https://www.gov.ca.gov/2025/08/07/governor-newsom-partners-with-worlds-leading-tech-companies-to-prepare-californians-for-ai-future/.

14 Mark Baldassare, Dean Bonner, Lauren Mora, and Deja Thomas, "PPIC Statewide Survey: Californians and Their Economic Well-Being," Public Policy Institute of California, December 2024, https://www.ppic.org/publication/ppic-statewide-survey-californians-and-their-economic-well-being-december-2024/.

15 Luona Lin and Kim Parker, "U.S. Workers Are More Worried Than Hopeful About Future AI Use in the Workplace," Pew Research Center, February 25, 2025, https://www.pewresearch.org/social-trends/2025/02/25/u-s-workers-are-more-worried-than-hopeful-about-future-ai-use-in-the-workplace/.

16 Molly Kinder, "Hollywood Writers Went on Strike to Protect their Livelihoods from Generative AI. Their Remarkable Victory Matters for All Workers," Brookings Institution, April 12, 2024, https://www.brookings.edu/articles/hollywood-writers-went-on-strike-to-protect-their-livelihoods-from-generative-ai-their-remarkable-victory-matters-for-all-workers/.

17 Brynjolfsson, Chandar, and Chen, "Canaries in the Coal Mine?"

18 "California Consumer Privacy Act of 2018," California Privacy Protection Agency, January 2025, https://cppa.ca.gov/regulations/pdf/ccpa_statute.pdf.

19 Scott Kohler, "State AI Regulation Survived a Federal Ban. What Comes Next?," Carnegie Endowment for International Peace, July 3, 2025, https://carnegieendowment.org/emissary/2025/07/ai-congress-bill-state-ban-what-next?lang=en.

20 "Our Approach to Data and AI," OpenAI, May 7, 2024, https://openai.com/index/approach-to-data-and-ai/.

21 Jesus Jiménez, "Worried About Meta Using Your Instagram to Train Its A.I.? Here's What to Know," *New York Times*, September 26, 2024, https://www.nytimes.com/article/meta-ai-scraping-policy.html.

22 Zach Vallese, "Creators Say They Didn't Know Google Uses YouTube to Train AI," CNBC, June 19, 2025, https://www.cnbc.com/2025/06/19/google-youtube-ai-training-veo-3.html.

23 Lila Shroff, "Shh, ChatGPT. That's a Secret.," *The Atlantic*, October 2, 2024, https://www.theatlantic.com/technology/archive/2024/10/chatbot-transcript-data-advertising/680112/.

24 "Our Approach to Data and AI," OpenAI, May 7, 2024, https://openai.com/index/approach-to-data-and-ai/.

25 Ian Klaus, Dan Hendrycks, Ketan Ramakrishnan, Ion Stoica, and Lauren Wagner, "The Future of AI Regulation: A California Bill Shaping the Debate," Carnegie Endowment for International Peace, panel discussion, September 12, 2024, https://carnegieendowment.org/events/2024/09/how-should-ai-be-regulated-a-california-bill-shaping-the-debate?lang=en.

26 Amy Jones, "Insights Into the Integration of AI in Government," Ernst & Young, September 18, 2024, https://www.ey.com/en_us/industries/government-public-sector/insights-into-the-integration-of-ai-in-government.

27 Holden Karnofsky, "If-Then Commitments for AI Risk Reduction," Carnegie Endowment for International Peace, September 13, 2024, https://carnegieendowment.org/research/2024/09/if-then-commitments-for-ai-risk-reduction?lang=en.

28 Holden Karnofsky, "A Sketch of Potential Tripwire Capabilities for AI," Carnegie Endowment for International Peace, December 10, 2024, https://carnegieendowment.org/research/2024/12/a-sketch-of-potential-tripwire-capabilities-for-ai?lang=en.

29 Daniel Privitera, Tamay Besiroglu, Rishi Bommasani, et al., "International AI Safety Report 2025," Department for Science, Innovation and Technology, February 18, 2025, https://www.gov.uk/government/publications/international-ai-safety-report-2025/international-ai-safety-report-2025.

30 Rishi Bommasani, Scott Singer, Ruth E. Appel, et al. "The California Report on Frontier AI Policy," Joint California Policy Working Group on AI Frontier Models, June 17, 2025, https://carnegieendowment.org/research/2025/06/the-california-report-on-frontier-ai-policy?lang=en.

31 Deep Ganguli, Danny Hernandez, Liane Lovitt, et al., "Predictability and Surprise in Large Generative Models," Anthropic, June 24, 2022, https://arxiv.org/pdf/2202.07785.

32 Holden Karnofsky, "AI Has Been Surprising for Years," Carnegie Endowment for International Peace, January 6, 2025, https://carnegieendowment.org/research/2025/01/ai-has-been-surprising-for-years?lang=en.

33 Point 7 of executive summary.

34 Rishi Bommasani, Scott Singer, Ruth E. Appel, et al. "The California Report on Frontier AI Policy," Joint California Policy Working Group on AI Frontier Models, June 17, 2025, https://carnegieendowment.org/research/2025/06/the-california-report-on-frontier-ai-policy?lang=en.

Carnegie Endowment for International Peace

In a complex, changing, and increasingly contested world, the Carnegie Endowment generates strategic ideas, supports diplomacy, and trains the next generation of international scholar-practitioners to help countries and institutions take on the most difficult global problems and advance peace. With a global network of more than 170 scholars across twenty countries, Carnegie is renowned for its independent analysis of major global problems and understanding of regional contexts.

Carnegie California

Carnegie California links developments in California and the West Coast with national and global conversations around technology, subnational affairs, and trans-Pacific relationships. At distance from national capitals, and located in one of the world's great experiments in pluralist democracy, Carnegie California engages a wide array of stakeholders as partners in its research and policy engagement.

CARNEGIE
ENDOWMENT FOR
INTERNATIONAL PEACE

CarnegieEndowment.org