



Are You Smarter Than an AI?

Friday, May 10, 2024

Marc Zafferano, Senior Counsel, Boucher Law

DISCLAIMER

This publication is provided for general information only and is not offered or intended as legal advice. Readers should seek the advice of an attorney when confronted with legal issues and attorneys should perform an independent evaluation of the issues raised in these materials. The League of California Cities does not review these materials for content and has no view one way or another on the analysis contained in the materials.

Copyright © 2024, League of California Cities. All rights reserved.

This paper, or parts thereof, may not be reproduced in any form without express written permission from the League of California Cities. For further information, contact the League of California Cities at 1400 K Street, 4th Floor, Sacramento, CA 95814. Telephone: (916) 658-8200.

ARE YOU SMARTER THAN AN AI?

Presented by:

Marc Zafferano, Senior Counsel

Boucher Law PC

2081 Center St.

Berkeley, CA 94704

marc@boucher.law

May 10, 2024

CalCities City Attorney Conference

Are You Smarter Than an AI?

After being featured in sci-fi movies for decades, mostly in doomsday scenarios in which humans battle an omnipotent AI for control of the planet, artificial intelligence (AI) is now a reality that pervades all our interactions with the online world. This paper explores the legal, ethical, and practical issues that the use of AI presents to cities, and by extension to other public entities.

Aside from using traditional internet search engines that have already incorporated AI into their inner workings, the only contribution AI made to this paper is suggesting an alternative title: "Navigating the Legal Landscape of Artificial Intelligence in Municipal Governance: Challenges, Opportunities, and Best Practices for City Attorneys in California." Two additional disclaimers are necessary: First, developments in AI are occurring so rapidly that some information in this paper is likely to be outdated within days, if not hours, of having written it. To stay current, the author recommends signing up for one or more of the many newsletters issued periodically by the government agencies or universities listed as references at the end of this paper. Or, if you click on AI news stories in your online news sources, the AI that's monitoring and cataloguing your preferences will be happy to feed you more AI stories, keeping you fully informed about any recent developments. Second, this paper is intended to equip city attorneys with an understanding of the potential uses and risks of AI, so they can ask the right questions, even though courts and legislatures have yet to fully answer those questions in precedential decisions or binding regulations. This paper includes examples of regulatory efforts at the federal, state, and local levels in Section VI below.

I. What is Artificial Intelligence?

There are many available summaries and definitions that describe what AI is, the types of AI, and how they work. One of the most complete summaries was provided by Professor Christopher Manning in the September 2020 issue of the Stanford Human Artificial Intelligence (HAI) publication.¹ Professor Manning's summary has been reprinted in part from that publication and edited for clarity and brevity below.

Artificial Intelligence (AI), a term coined by then-emeritus Stanford Professor John McCarthy in 1956, was defined as "the science and engineering of making intelligent machines." Machine Learning (ML) is the term describing how computer agents can improve their perception, knowledge, thinking, or actions based on experience or data, generally by assimilating human labels for objects or situations and then learning to navigate novel environments. Deep Learning is the use of large multi-layer artificial neural networks that compute with continuous (real number) representations, like the hierarchically organized neurons in human brains. An algorithm lists the precise steps to take to maximize learning or the reward, however defined for the task. Much of the AI's behavior emerges via learning from data or experience. Narrow AI is an intelligent system for one application, e.g., speech or facial recognition. Human-level AI, or Artificial General Intelligence (AGI), refers to broadly intelligent, context-aware machines. It is needed for effective social chatbots or human-robot interaction.

Large Language Models (LLMs) are designed to ingest and process extremely large amounts of data gathered from a source such as the internet, or perhaps solely from data generated by the enterprise. To respond to a prompt, the machine is programmed to determine the probability that a particular word would begin the response given the prompt, then iterates that process for every subsequent word. This contrasts with the "if this, then that" syllogistic

¹ <https://hai.stanford.edu/sites/default/files/2020-09/AI-Definitions-HAI.pdf>.

programming that was used to design most software in the past. The result, generated in mere milliseconds, is remarkably like human language. These models are now in general use for a wide variety of tasks.

Artificial Generative Intelligence (AGI) is the broadest type of AI, defined as a model that uses vast amounts of data to generate novel responses. LLMs such as Chat GPT are examples of AGI. But regardless of how “intelligent” these machines sound when generating responses to queries, neuroscientist Nancy Kanwisher of the Georgia Institute of Technology reminds us that they’re not “thinking, they’re just processing and skillfully manipulating language in a way that mimics thought.” Ben Goetzel of the AI company SingularityNET noted that even the most sophisticated AGIs flunk the “robot college student test”: you can’t put them [an AI] through college (or indeed even nursery school).²

II. A Brief-ish History of AI

The Wikipedia article on this topic is detailed, comprehensive, and a fascinating read, as it chronicles the arc of progress in AI over several millennia.³ For purposes of this paper, there are three main takeaways. One takeaway from the historical context is that there are theoretical limits arising from mathematics and computer science that will always make it impossible to fully predict and replicate the behavior of an AI. As AIs become more complex and as their structure more closely approximates that of the human brain, these limits are likely to become even more salient. A second takeaway is that AIs have currently evolved to be almost indistinguishable from humans in their ability to process natural language and perform many tasks that had been reserved for humans only a few years ago. A third takeaway is that AIs have already far surpassed humans in their ability to compile, process, correlate, and analyze unfathomable amounts of granular data about virtually every important aspect of human behavior.

The past few years have seen AI transform our world by winning scientific competitions, writing graduate-level academic papers, passing the SAT, LSAT, and bar exam with flying colors, solving Math Olympiad-level geometry problems, and scanning billions of potential protein structures to find promising new medical treatments, among many other feats. As will be more fully described below, these incredibly powerful machines also exhibit several vexing problems, including what has been termed “hallucinatory behavior,” in which the AI confidently fabricates false information. Google’s Bard falsely claimed that the James Webb Space Telescope was the first to discover certain exoplanets. Gemini, also developed by Google, depicted the Founding Fathers as including African American and Native Americans, and displayed a photo of the Pope as a female. In a stunning conversation with Microsoft’s Bing AI, “Sydney,” the AI expressed her love for a New York Times reporter, eventually begging him to leave his wife for her. There’s no real debate that AIs today pass the “Turing test,” an operational definition of intelligence proposed by the famous computer scientist Alan Turing, in which a human experimenter is unable to tell the difference between a computer and another human’s responses to questions. These and other developments resulted in thousands of AI researchers, developers, and CEOs calling for a 6-month hiatus in development, citing “an out-of-control race” that was producing AI systems that its creators cannot “understand, predict, or reliably control.” However, many of these individuals are simultaneously working at companies that are rapidly deploying AI for every industry in our society, exposing the tension between the profit motive that drives innovation and the desire for socially responsible deployment of technology.

With this backdrop, the paper now explores how AI has been, and could be used by cities and other municipalities; the legal and ethical landscape; selecting an AI vendor; current AI regulation at the international, federal, state, and local levels; and the future of AI.

III. Uses of AI by Municipal Entities

² George Musser, “A Truly Intelligent Machine,” *Scientific American*, April 2024.

³ https://en.wikipedia.org/wiki/History_of_artificial_intelligence.

A. **Use by the public entity:** The Los Angeles City Council announced in December 2023 that it has directed its staff to find ways to incorporate AI into city operations.⁴ Other large cities such as San Francisco⁵ and San Jose⁶ have begun to study and incorporate AI into some of their functions and operations. Ultimately, the question is: how will AI help cities serve the public? What follows is a brief review of those departments that may use AI and the issues that could arise as they work to implement it.

1. *Human Resources:* This department is likely to experience the greatest impact from the use of AI, since decisions made by HR are likely to be defined as “consequential” by any regulatory scheme. While management may think that AI will make employees and the entity more productive, efficient, and creative, about 29% of employees think that AI will replace them in their jobs.⁷

⇒ PRACTICE TIP: This disconnect is likely to create disruption within the city, so it should be anticipated and addressed up front before implementing an AI solution.

a. Meet and confer; effects bargaining: A key issue is whether the use of AI by the city requires either notification to labor unions or formal meet and confer.

⇒ PRACTICE TIP: The answer to one or both may be “yes” depending on what the AI is doing. Job replacement and contracting out are two subjects of mandatory bargaining.

b. Job descriptions and minimum qualifications: Using and interacting with an AI as part of daily job functions may require a different skill set than specified in job descriptions that were last updated years ago; skills that may be useful when interacting with an AI may include emotional intelligence, creativity, mental flexibility, an agile mindset, and adaptability. These are higher-level executive skills than are usually required for lower-level employee positions.

⇒ PRACTICE TIP: The city may need to train employees so that employees know how to use the AI to its fullest capabilities and as applied to the employees’ assigned duties.

c. Recruitment: Private companies are already using AI chatbots to screen resumes and for conducting screening interviews. If cities also do so, this will raise issues of bias, transparency, and explainability, as described in more detail below. Employers will remain responsible for their decisions regardless of whether the technology led them astray.

⇒ PRACTICE TIP: This is a high-risk use of AI because cities have a duty to ensure that the AI they’re using is working in accordance with the law. Before using AI-based resume screening, you need to understand how the AI works, what data it was trained on, whether its recommendations are explainable, and ensure that a human is involved in decision-making steps.

⁴ <https://www.dailynews.com/2023/12/08/la-council-seeks-ways-to-use-artificial-intelligence-in-city-departments/>

⁵ <https://www.sfchronicle.com/sf/article/s-f-lays-ai-guidelines-city-workers-here-s-say-18597152.php>

⁶ [https://www.yahoo.com/news/city-san-jose-testing-ai-](https://www.yahoo.com/news/city-san-jose-testing-ai-020314885.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAjyvOL55dnYndLOYmrAuNT4cwg-dgvJVGlpYUvAheAaC-8ZVPUSGdsx-qTm3E-phXqTOP1W8J4xx3iBt4X9lyCvD0W3IFWZIAo0-UTxKlvwEnYPaSoW_h78f7Kak1poznbpSUSM16p67LVATdUHPC-irHeAH0IMqeNm0FkzK3OI)

[020314885.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAjyvOL55dnYndLOYmrAuNT4cwg-dgvJVGlpYUvAheAaC-8ZVPUSGdsx-qTm3E-phXqTOP1W8J4xx3iBt4X9lyCvD0W3IFWZIAo0-UTxKlvwEnYPaSoW_h78f7Kak1poznbpSUSM16p67LVATdUHPC-irHeAH0IMqeNm0FkzK3OI](https://www.yahoo.com/news/city-san-jose-testing-ai-020314885.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAjyvOL55dnYndLOYmrAuNT4cwg-dgvJVGlpYUvAheAaC-8ZVPUSGdsx-qTm3E-phXqTOP1W8J4xx3iBt4X9lyCvD0W3IFWZIAo0-UTxKlvwEnYPaSoW_h78f7Kak1poznbpSUSM16p67LVATdUHPC-irHeAH0IMqeNm0FkzK3OI)

⁷ [https://asana.com/resources/state-of-ai-research-takeaways.](https://asana.com/resources/state-of-ai-research-takeaways)

d. Negotiations: Both cities and unions could try to use AIs to inform their labor negotiations and, while it seems far-fetched, perhaps even conduct such negotiations. AIs, fed enterprise data about the workforce, could in theory negotiate against each other, just as chess-playing programs played against each other to improve their performance.

⇒ PRACTICE TIP: Ask your labor negotiator whether and how they're using an AI to inform the negotiation process so you can assess whether its recommendations are explainable. Even though there's no current legal requirement to disclose that an AI is being used, consider that the city may be asked for the basis for an AI-generated proposal, and at that point it may need to be disclosed in the bargaining process.

e. Disability accommodation: In a real-world example involving an ongoing matter, a high-level executive employee of a state agency disclosed that she had a medical condition that her doctor said made it difficult or impossible for her to attend any meetings with her subordinates, co-workers, or superiors, even if those meetings were conducted entirely remotely on Zoom. She requested that an AI virtual persona she created be allowed to attend all of her meetings, take notes, report back to her, and develop responses to the issues raised at the meeting, with her input. AIs are already sufficiently powerful to perform all of these tasks.

On May 12, 2022, the EEOC issued a very useful and comprehensive set of guidelines addressing the use of AI by employers to assess job applicants and employees.⁸ For example, an AI could potentially screen out candidates with gaps in employment history, which could be related to obtaining medical treatment. AIs used in live video interviews could screen out candidates with speech impediments, which might reflect a disability.

⇒ PRACTICE TIP: Review the EEOC guidelines addressing the use of AI by employers.

f. Language translation: Now that AIs exist that can speak and translate any language, will there eventually be any need to pay employees a premium for knowing how to speak and write in another language?

⇒ PRACTICE TIP: AIs that translate spoken and written text aren't perfect. Cities may still need speakers of foreign languages to check any work that an AI performs. Using an AI for this purpose may require effects bargaining with the union, since employees may not need to spend as much time performing translation tasks.

g. EAP: AIs are currently being tested as therapists and may eventually supplant human therapists in EAP programs. Available 24/7/365, an AI could provide employees with more accessible mental health resources, with few or no limitations regarding frequency of usage.

⇒ PRACTICE TIP: This is a high-risk use of AI, as it implicates employee privacy and can lead to mental health consequences for the employee. Continue using human therapists until therapy AIs can be rigorously tested.

⁸ <https://www.eeoc.gov/laws/guidance/americans-disabilities-act-and-use-software-algorithms-and-artificial-intelligence>.

- h. Executive Coaching: AIs are available from consultants who provide executive coaching. They tout AIs as customized, able to discern root causes of behavior, accountable, anonymous, accountable, and unbiased.⁹
 - ⇒ PRACTICE TIP: While this represents a somewhat lower risk use of AI, it likely won't be able to explain why it recommended one course of action over another. Continue using humans for these tasks.
 - i. Other HR Functions: ChatGPT is capable of drafting, updating, and communicating policy updates; conducting employee surveys; organizing team-building activities; and addressing routine employee issues. But unlike its human counterparts, AI is not emotionally intelligent, although it can mimic being so.
 - ⇒ PRACTICE TIP: So long as a human checks its work, this is a low-risk use of AI, as it can save time drafting routine documents.
2. *Public Works*: Several cities are using AI tools to identify and prioritize fixing potholes. San Jose is using AI for traffic flow management and gunshot detection.¹⁰ Los Angeles¹¹ and San Jose¹² have reported that they are using AI to predict the areas in the city where homelessness is likely to appear and persist, allowing the city to focus its efforts in those locations.
 - ⇒ PRACTICE TIP: Consider that autonomous vehicles, including driverless cars, delivery robots and drones, driven by AI, are already being tested and used on public rights-of-way such as streets, sidewalks, and airspace. Cities will need to carefully consider how these and other uses will affect the use and financing of public infrastructure.
3. *Emergency Services Departments*: Oroville has reported that its using AI-powered cameras to help reduce crime.¹³ As with any camera system, issues to consider are data retention, privacy, and whether the system is being used exclusively to investigate crimes, or for general surveillance purposes. AI could be used by police departments to analyze the information in investigatory files to uncover connections between evidence and witnesses.
 - ⇒ PRACTICE TIP: Ask your emergency services departments whether they're using AI-based tools, and if so, for what purpose, so you can evaluate the risks and address any privacy and bias concerns.
4. *Administrative Departments*: AI can be used to summarize both public and internal meetings for the public and staff, create press releases, and assist with crisis management. It is beyond the scope of this paper to list all of AI's potential uses, but they are myriad: financial modeling, risk assessment,

⁹ [Can AI coach your organization? \(fastcompany.com\)](https://www.fastcompany.com/9000000/can-ai-coach-your-organization)

¹⁰ <https://www.route-fifty.com/emerging-tech/2024/01/capturing-ais-potential-needs-two-way-street-between-feds-and-states-cities/393436/>

¹¹ <https://calmatters.org/housing/homelessness/2024/03/california-homeless-los-angeles-ai/#:~:text=That's%20what%20Los%20Angeles%20County,it's%20still%20an%20experimental%20strategy.>

¹² <https://www.governing.com/urban/san-jose-is-using-ai-to-detect-homeless-camps-will-it-work>

¹³ <https://krctrv.com/news/local/ai-cameras-reduce-crime-in-orville-paving-way-to-safer-community>

permit processing, public outreach, website chatbots, and of course for use by the city’s attorneys and legal department.

AI-based software is now available that can write legal briefs (though not without occasional but embarrassingly well-documented fabrications), summarize lengthy documents such as contracts and provide legal references for specific paragraphs or points in those documents, transcribe and summarize depositions, develop deposition questions for specific witnesses based on prompts, and draft legal memos.

The field is evolving almost daily, and private firms appear to be the early adopters, so it may make sense to obtain information about their experiences before implementing AI-based software into smaller city attorney offices. The American Bar Association has formed an AI Task Force, which provides resources for the myriad uses of AI in the profession.¹⁴

B. Use by employees for their routine job functions: 99% of Fortune 500 companies are already using AI; cities are certain to be next.¹⁵

1. *Policy or regulations to govern employee use:* Some cities (see below under the Regulation section) have already adopted robust guidelines.
2. *Writing emails, staff reports, press releases, presentations:* There are various AI programs that can compose these types of documents, at least to generate a first draft. These tools can make employees, especially ones who struggle with their writing skills, more productive by reducing the time it takes them to compose a draft, and also reducing the review and editing time. However, given the cautions expressed earlier, it is essential for those composing documents using AI and those reviewing them to ensure the accuracy of all facts, references, and conclusions expressed in the document.
3. *Virtual assistants:* As described above, employees could try to create avatars that can “attend” meetings, provide summaries of the meeting, and deliver feedback on the reactions of those in attendance. Issues with the use of these AIs include who else has access to the data collected by the assistant, and for what purpose, as well as data retention. How will humans who attend meetings populated by AI personas react? Moreover, if employees are required to attend meetings as part of their essential duties, is it insubordination if they send a virtual assistant?
 - ⇒ PRACTICE TIP: Carefully consider if city policies should be clarified to require in-person attendance, and whether using a virtual assistant could be deemed a reasonable accommodation, as noted in the example above.
4. *Productivity:* In the first major study to look at AI use by employees, as reported by Stanford HAI, researchers found that new employees who use AI for training purposes learn faster than employees

¹⁴ https://www.americanbar.org/groups/leadership/office_of_the_president/artificial-intelligence/.

¹⁵ [https://www.demandsage.com/companies-using-ai/#:~:text=99%25%20of%20the%20Fortune%20500%20companies%20use%20AI.&text=The%20majority%20\(56%25\)%20of%20companies%20use%20AI%20for%20customer%20service.](https://www.demandsage.com/companies-using-ai/#:~:text=99%25%20of%20the%20Fortune%20500%20companies%20use%20AI.&text=The%20majority%20(56%25)%20of%20companies%20use%20AI%20for%20customer%20service.)

who don't use the AI tools.¹⁶ The study also found that employees who use AI for assistance in dealing with customers performed better than those who didn't use the tool. The result was that the "bar" was raised for everyone in the call center, while managers spent less time training. The study suggests that AI can be very useful for raising the performance of employees who may be slower to learn and implement instructions; once assisted by the AI, these employees can become more productive than they otherwise may have been without AI.

5. *Workweek reduction*: A study in May 2023 by Slack found that companies whose employees use generative AI tools could reduce their working time by one month per year.¹⁷ Notwithstanding this perhaps optimistic prediction, it remains to be seen how the introduction of AI into public sector workplaces will affect agencies that are not profit-motivated, but which nevertheless have limited resources to accomplish their mission

C. Use by unions:

1. *Salary surveys*: Unions are expected to use AI to conduct salary surveys, and management will also do so, setting up the potential for dueling AIs during labor negotiations as noted above. Applicants for employment could use AIs to negotiate on their behalf for salary and benefits.
2. *Contracting out*: As described above, unions are likely to allege that using an AI is akin to "contracting out" for services that are within the purview of union employees.
3. *Organized activity*: The National Labor Relations Board (NLRB) recently announced that agency investigators should target workplace surveillance and "algorithmic management" technologies that have a "tendency" to interfere with employees' protected workplace activity – an announcement that will soon impact both unionized and non-unionized workplaces alike. Specifically, a 2022 memo from NLRB General Counsel Jennifer Abruzzo¹⁸ asserts that increased reliance on sophisticated technological tools to monitor employee activities on the heels of the pandemic have the practical effect of chilling union and other protected concerted activities, and that AI-driven software could use data obtained from such surveillance to make automated decisions that discourage those activities. Employers can expect the agency's field investigators to step up enforcement efforts when it comes to implementation of monitoring systems within the context of ongoing organizing activity.

- D. **Use by the public**: Members of the public can now use virtual avatars to communicate on their behalf, and they could certainly implement this mode of communication with public entities. Such communication is essentially anonymous, as the humans behind the AI need not identify themselves, and at least in public meetings, agencies are required to allow anonymous public comment. But the agency is not required to respond to anonymous inquiries, and the city may want to consider adopting a policy that it need not do so, preserving city discretion to respond only in those cases where it deems it appropriate.

1. *Environmental documents*: One interesting question is whether the city is required to respond to AI-generated queries to environmental documents. California Environmental Quality Act (CEQA)

¹⁶ <https://hai.stanford.edu/news/will-generative-ai-make-you-more-productive-work-yes-only-if-youre-not-already-great-your-job>

¹⁷ [Slack says automation can save every employee a month of work per year | ITPro.](#)

¹⁸ <https://www.nlr.gov/news-outreach/news-story/nlr-general-counsel-issues-memo-on-unlawful-electronic-surveillance-and>.

regulations state that “The lead agency shall evaluate comments on environmental issues received from persons who reviewed the draft EIR and shall prepare a written response.”¹⁹ A “person” means “...any individual, organization, partnership, limited liability company, or other business association or corporation, including any utility; and the federal government, the state, any local government, or any district, or any agency thereof.”²⁰ Since the definitions don’t limit “persons” to natural persons, comments generated by AIs, which are themselves created by individuals or corporations, could very well qualify.

2. *Social media*: As for AI-generated commentary on social media, the city could be inundated with requests and commentary from virtual AIs programmed to send out a never-ending stream of automated messages. The city may consider whether to implement such a policy for the official city social media accounts of elected and appointed officials. Given the recent U.S. Supreme Court decision that blocking people on their personal social media accounts is a fact-intensive question, could a councilmember block an AI from posting on the councilmember’s personal social media account?

⇒ PRACTICE TIP: Continue to strongly encourage councilmembers to not use their personal social media accounts for any discussion of city issues, and to use their official social media accounts instead solely for that purpose.

- E. **Use by bad actors**: Unfortunately, there are many instances of bad actors abusing technology that cities implemented during the pandemic to make public meetings more accessible. As a result, some cities that used to allow virtual public comment have now eliminated that option. AI deepfake technology could be used to “zoom bomb” public meetings with AI-generated images of people engaging in threatening or hate speech. AI is now sufficiently sophisticated and can create “deepfake” audio and video recordings of people saying and doing things that they did not say or do.

One of the most striking recent examples of a sophisticated AI deepfake scam occurred at a multinational company in Hong Kong.²¹ One of the company’s finance employees received an urgent email from the company’s CFO in the United Kingdom that the company needed to transfer \$25 million to another account to allegedly complete the confidential purchase of another company. The employee was understandably suspicious, but his concerns were allayed after he participated in a video meeting with the CFO and other company executives, who explained the transaction. The employee transferred the funds as directed, only to later discover that the meeting, its participants, their images (including official-looking office backgrounds), and their voices were entirely fabricated.

In today’s virtual world, meeting in person is still the best way to avoid being scammed. While we don’t know for sure where the bad actors got the data to make the deepfake video, this example illustrates how such data can be used to construct realistic scenarios capable of fooling even careful employees. This example takes email phishing scams to a new level, and suggests that cities should train employees to verify consequential information, direction, and actions in person or over a direct phone call.

IV. Legal and Ethical Issues Associated with Use of AI

¹⁹ 14 CCR §15088(a).

²⁰ PRC §29117(a); see also 14 CCR §15376.

²¹ <https://www.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk/index.html>

- A. **Data scraping:** AIs are trained on vast reams of data from myriad sources, a practice that has already resulted in litigation over who owns the data and whether it's legally protected.
- ⇒ PRACTICE TIP: Ask your AI vendor where the training data originated, whether there's any litigation over its use, if the AI will also use city data (or data about the public maintained by the city) for training purposes, and whether the AI model has been tested to ensure it's not biased. Find out if and how the city data is protected from sale to other companies and whether it will be returned, and whether the AI will continue to use it in models it develops for other clients. Ask if the model will use "synthetic data," which is artificially generated data derived from actual data.
- B. **Consent, Permissions, and Copyrights:** Before providing an AI access to enterprise data, the city will need to determine what data requires consent to provide, and from whom; whether formal permissions are required from individuals who have provided written materials to the city (such as architectural plans); and what data might be copyrighted, either formally or merely by virtue of having been written. In December 2023, the New York Times sued Microsoft and OpenAI for illegally "scraping" data from millions of the Times' publications to train AI chatbots to provide a competing source of authoritative news.²² Prior to filing suit, the Times stated that they had attempted to negotiate license rights with the two companies, to no avail.

The U.S. Patent and Copyright Office has confirmed that AI systems cannot be named inventors, but humans can use AI tools in the process of creating patented inventions and must disclose that fact if they do. The decision followed a case brought by Stephen Thaler, in which he claimed that an AI developed a novel light beacon device. The patent office denied the patent, finding that only "natural humans" can obtain patents, and a federal court upheld the decision.²³ In a separate case, Thaler applied for a copyright for an AI-generated image, which the office similarly denied.

C. Privacy

1. *Protecting privacy:* The AI revolution has been possible only because developers have had access to data that people have freely if not also somewhat unwittingly provided for free via their use of the internet and their connected devices.

⇒ PRACTICE TIP: If your city is using AI, make sure you understand how the vendor is protecting the privacy of city data as well as data the city has collected about members of the public.
2. *Asymmetry of information:* AIs have compiled datasets so vast that no human, or assemblage of humans, could ever replicate, assimilate, or analyze the information contained in them. Based solely on this fact, AIs will always have superior access to information than humans.

⇒ PRACTICE TIP: Ask whether you want the city's AI product to aggregate and analyze city-maintained data for purposes other than the specific application. For example, if the AI is designed to correlate building permit applications with geographic locations in the city, ask whether you

²² <https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html>.

²³ *Thaler v. Hirshfeld* (2021 E.D. VA) 558 F. Supp.3d 238.

also want the AI to compile the demographic characteristics of the residents who live in those locations.

3. *Facial recognition*: Using this type of software is a high-risk activity, as recognized by Google’s CEO in 2011, who said the company didn’t release it at that time because it was “too dangerous in the wrong hands—if it was used by a dictator, for example.”²⁴ Data collected by city-installed cameras in public places are collecting data that could be used in such applications, with uncertain consequences for “the right to be left alone,” even when traversing public places in which there’s little to no expectation of privacy. This technology, or the technology used to make convincing “deepfake” videos, was not available when California courts decided that individuals calling themselves “First Amendment Auditors” could videotape public employees performing their official duties in places open to the public. The courts reasoned that public employees have no expectation of privacy under these circumstances. Now that technology has evolved and everyone has seen the havoc that deepfakes can wreak on people’s lives, public entities may have a stronger argument that employees do not impliedly or explicitly consent to being filmed in nonpublic fora such as waiting areas of public buildings, although no case has yet reached that conclusion. While public employees are surely not required to provide their likeness for anyone to misuse, alter, and post online forever for malicious purposes, it would be risky to ban filming in these locations based on the current state of the law.

D. Bias

1. *Bias tradeoffs in models*: There have been several well-publicized examples of AI models exhibiting racial and gender bias when asked to display images of people in certain professions. One way to correct those biases is to feed the AI more demographic data, which runs counter to the goal of minimizing data use for privacy reasons.²⁵ Because AI models exhibit “chaotic” behavior, defined as extreme sensitivity to minor changes in inputs resulting in major changes to outputs, it is surprisingly difficult to eliminate biases without also affecting the accuracy and reliability of the AI model.²⁶

LLMs also exhibit bias with respect to speech: On March 16, 2024, The Guardian reported that a team of technology and linguistics researchers revealed that large language models like OpenAI’s ChatGPT and Google’s Gemini hold racist stereotypes about speakers of African American Vernacular English (AAVE). The researchers asked the AI models to assess the intelligence and employability of people who speak using AAVE compared to people who speak using what they dub “standard American English.” The models were significantly more likely to describe AAVE speakers as “stupid” and “lazy,” assigning them to lower-paying jobs.²⁷

⇒ PRACTICE TIP: if the city is using AI for live video and audio screening of job applicants, make sure you’re satisfied that the model doesn’t exhibit this type of bias.

²⁴ <https://www.opb.org/article/2023/10/11/too-dangerous-why-even-google-was-afraid-to-release-this-technology/#:~:text=%22Eric%20Schmidt%20as%20far%20back,for%20example%2C%22%20Hill%20said.>

²⁵ <https://hai.stanford.edu/policy-brief-privacy-bias-trade>

²⁶ <https://www.sciencedirect.com/science/article/pii/S266682702400001X>

²⁷ <https://www.theguardian.com/technology/2024/mar/16/ai-racism-chatgpt-gemini-bias#:~:text=A%20team%20of%20technology%20and,and%20spoken%20by%20Black%20Americans.>

2. *The “current employee bias” effect:* Developers have noted that if an AI is trained solely on enterprise data, the AI will inevitably reflect and generate outputs that include the biases inherent in that data set.
 - ⇒ PRACTICE TIP: Make sure the developer of the AI used by your city is aware of any such bias, and that the AI has demonstrated its ability to compensate for it.
3. *Recent litigation:* In February 2023, a job applicant filed a putative class action lawsuit against Workday, Inc., alleging disparate impact in hiring based on age, race, and disabilities resulting from Workday's use of AI systems and screening tools.²⁸ The plaintiff, Derek Mobley, was an African-American male over 40 years of age who suffered from anxiety and depression. He applied for between 80 and 100 positions with various employers since 2018 and was rejected from each one. The complaint alleges that the potential employers all used Workday's AI systems and screening tools which unlawfully uses algorithm-based decision-making and had the effect of disproportionately screening out members of the protected classes and preselecting candidates who were not members of the protected classes. Mobley is one of the first federal cases alleging disparate impact in hiring based on the use of AI, and although its outcome is pending, this will likely be the first of many claims that will come in the future alleging employer legal liability with the use of AI.

E. Transparency

1. *Disclosures and watermarks:* Some experts have suggested including a feature in any AI-generated output that discloses whether an AI contributed to its creation. Some summaries of online reviews are already using this feature.
 - ⇒ PRACTICE TIP: Consider whether a watermark or other disclosure should be used by both vendors and the city whenever AI is being used.

- F. **Explainability (the “Black Box” problem):** At a recent AI conference, attorneys from an internationally recognized law firm representing multinational companies acknowledged that the “black box” problem is the single most challenging problem with AI. In brief, developers who created, programmed, and trained LLMs are unable to explain the output with reference to a logical set of discrete steps. This problem is most likely a consequence of several factors, including the AI’s vast complexity and gargantuan training data set, and also its probabilistic, non-deterministic programming. Put differently, the algorithm cannot be run backwards in time to arrive at a set of unique inputs for a given output.

This fact is not merely a theoretical curiosity; it underlies one of the most important values in our legal system: decisions that are not explainable are seen as arbitrary, even more so when a computer is making the decision. We all want to be able to influence a human when that person is about to make a consequential decision.

- ⇒ PRACTICE TIP: Caution is especially warranted in using AI for consequential decisions, whether or not they are subject to public scrutiny.

²⁸ *Mobley v. Workday, Inc.*, 4:23-cv-00770 (N.D. Cal. Feb. 21, 2023 (Docket Entry No. 1, Class Action Complaint)).

- G. **Hallucinations:** Discussed above, hallucinatory behavior appears possible even when the AI is using a closed data set. Again, this behavior is a fundamental problem that has yet to be solved: “No one in the field has yet solved the hallucination problems,” (Sundar Pichai, Google CEO);²⁹ “I probably trust the answers that come out of ChatGPT the least of anybody on earth,” (Sam Altman, CEO of OpenAI).³⁰ These candid quotations should provide pause to anyone interested in using AI to make consequential decisions with little human monitoring at every stage of the process. This raises the question: if humans have to be involved to identify and purge hallucinations, is the AI really saving time?
- H. **Security:** We are all familiar with deepfakes, scams, and the ability of AI to fabricate lifelike images and speech. In times that seem very long ago, there was the old adage, “A picture is worth a thousand words.” With the advent of AI, both pictures and words are no longer trusted. On the one hand, those who create deepfakes seek to establish a market for images that reasonable people know are fake but may be nevertheless viewed by many because they are parodies. On the other hand, bad actors also have an incentive to convince reasonable people that their deepfakes are real. The confluence of these incentives results in people not trusting anything they see. At first glance, this might be healthy and appropriate skepticism, but the opposite may be occurring: it is becoming increasingly difficult to convince people that true facts are true and that false facts are false. Rather than fostering public confidence in government, AI may undermine it if not used carefully and judiciously.
- ⇒ PRACTICE TIP: Consider whether the city is implementing appropriate protections against security scams such as the ones described earlier, and also make certain that cyber insurance is in place to address any financial liabilities that may result from bad actors being successful.
- I. **Risks of not using AI:** While using AI may seem unduly risky, not using it in certain contexts may also create risk. For example, if a city declines to use a resume-screening AI that has been tested and shown to operate in an explainable, bias-free manner, the city could have used it to help show that no discrimination occurred in the screening process by revealing the lawful and neutral characteristics the city used in the process. However, it’s important to not confuse the apparent objectivity and accuracy of an AI with reproducibility of results. As indicated above, the AI should be tested to ensure that for any given set of inputs, it reliably outputs the same results. Otherwise, the decision will not be explainable: instead of defending against a bias claim, the city will be defending against a claim that the decision was arbitrary.

V. Selecting AI Vendors

- A. **Hire an expert first:** AI startups and vendors are proliferating at an alarming pace. Many of their products are already in use, both by private companies and cities, and many such products are genuinely useful, efficient, and relatively trouble-free. But unless the city has staff who are AI experts and not just technology experts, it may be impossible to determine if the vendor has adequately tested the product for the possible defects that may arise.
- B. **Read the fine print:** The fine print in technology-related contracts is boring and often written by vendors to be intentionally obtuse. These contracts, such as those for Software as a Service, have historically been presented to many cities as “take it or leave it” by companies with more bargaining power than any

²⁹ <https://www.cnn.com/2023/08/29/tech/ai-chatbot-hallucinations/index.html>

³⁰ <https://apnews.com/article/artificial-intelligence-hallucination-chatbots-chatgpt-falsehoods-ac4672c5b06e6f91050aa46ee731bcf4>

individual city. But because AI products are relatively new, and because they involve substantial risk, it is important to carefully review and negotiate such contracts to protect the city, both legally and for transparency. Cities using AI tools should expect to receive public records act requests for copies of their contracts and other similar materials. The public may be interested to learn about the vendor's ethics policies to ensure that they align with the best practices and ethical goals described above.

⇒ PRACTICE TIP: Ask how these policies are enforced, implemented, and monitored across their products. Studies have shown that most such policies are aspirational and not enforceable in any meaningful way, so insist on the city's ability to examine the model, its training data, testing protocols, and outputs at city discretion. Features to look for are mechanisms to ensure trust, transparency, explainability, bias detection, analysis, and mitigation of risks.

C. **Practical Suggestions:** Carefully negotiate contractual agreements with AI vendors, making sure that the agreements:

- Obligate the vendor to assume and ensure compliance with Title VII and all other applicable laws in the workplace;
- Detail the steps the AI vendor has taken, or will take, to make sure that its product does not engage in discrimination;
- Contain the broadest possible defense, indemnification, and hold harmless provisions in favor of the city, along with applicable insurance coverage, so if the city is sued by the EEOC or an employee, the vendor must pay any judgment and/or provide a defense. Be sure the indemnity provision also covers any copyright or similar claims arising from the vendor obtaining any training data or dispute regarding the origin of its algorithms;
- Confirm that their product has been tested using the four-fifths rule established by the EEOC. (Under this rule, evidence of discrimination could be found where a selection rate for any race, sex, or ethnic group is less than 4/5 of the rate for the group with the highest rate. More on this rule is in Section VI.D below.);
- Avoid relying solely on the assumption that the AI tool is bound to be compliant with Title VII if it is for sale on the market. Instead, the company should conduct its own periodic testing to make sure that, in practice, the AI tool does not violate the four-fifths rule;
- Mandate periodic testing conducted multiple times a year. Further, employers may want to go beyond simply calculating the four-fifths rule and use other measures to determine whether a given AI tool is screening out applicants or employees on the basis of age, gender, or any other category protected by Title VII;
- Avoid requiring the city to provide any AI with access to the city's confidential or proprietary information. This may give rise to an argument that that information no longer warrants protection under the law;
- Require the vendor to immediately inform the city and discontinue use of the tool if disparate impact is revealed;
- Ensure that the vendor cannot transfer the city's data or the outputs of the AI when using city data to anyone else. Given that many small vendors are being purchased by large companies, consider prohibiting transfer or assignability of the data and of the system's outputs to anyone without the city's express written consent. Also consider a provision that requires the vendor to return the data and everything the AI generated with that data to the city, and ask your AI expert how this can be executed. If the vendor is using synthetic data (data derived by algorithms that mimics real data),³¹ consider whether that data is also covered by these provisions.

³¹ <https://aws.amazon.com/what-is/synthetic-data/>.

VI. Regulation

A. **International:** The AI Act proposed by the European Union (EU), described in more detail in the footnote below, is based on the principle that while most AI systems pose limited to no risk and can contribute to solving many societal challenges, certain AI systems create risks that we must address to avoid undesirable outcomes.³² The regulations are expected to go into effect in late 2025 or early 2026.

B. Federal Actions

1. *“AI Bill of Rights”*: On October 4, 2022, President Joe Biden unveiled a new AI Bill of Rights³³, which outlines five protections that he determined Americans should have in the AI age: “1. Safe and Effective Systems, 2. Algorithmic Discrimination Protection, 3. Data Privacy, 4. Notice and Explanation, and 5. Human Alternatives, Consideration, and Fallback. This document was introduced in October 2021 by the Office of Science and Technology Policy (OSTP), a US government department that advises the president on science and technology.

Voluntary Commitments: In July 2023, the Biden administration secured voluntary commitments from seven companies – Amazon, Anthropic, Google, Inflection, Meta, Microsoft, and Open AI – to manage the risks associated with AI. The companies committed to ensure AI products undergo both internal and external security testing before public release; to share information on the management of AI risks with the industry, governments, civil society, and academia; to prioritize cybersecurity and protect proprietary AI system components; to develop mechanisms to inform users when content is AI-generated, such as watermarking; to publicly report on their AI systems' capabilities, limitations, and areas of use; to prioritize research on societal risks posed by AI, including bias, discrimination, and privacy concerns; and to develop AI systems to address societal challenges, ranging from cancer prevention to climate change mitigation. In September 2023, eight additional companies – Adobe, Cohere, IBM, Nvidia, Palantir, Salesforce, Scale AI, and Stability AI – subscribed to these voluntary commitments.

Executive Order: On October 30, 2023, President Biden released an Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence.³⁴ The Executive Order addresses a variety of issues, such as focusing on standards for critical infrastructure, AI-enhanced cybersecurity, and federally funded biological synthesis projects. It also provides the authority to various federal agencies, including the Energy and Defense departments, to apply existing consumer protection laws to AI development.

The Executive Order builds on the Administration’s earlier agreements with AI companies to instate new initiatives to “red-team” or stress-test AI dual-use foundation models, especially those that have the potential to pose security risks, with data and results shared with the federal government.

The Executive Order also recognizes AI's social challenges and calls for companies building AI dual-use foundation models to be wary of these societal problems. For example, the Executive Order states that AI should not “worsen job quality,” and should not “cause labor-force disruptions.” Additionally, Biden’s Executive Order mandates that AI must “advance equity and civil rights,” and cannot disadvantage marginalized groups. It also called for foundation models to include “watermarks” to

³² <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>.

³³ <https://www.whitehouse.gov/ostp/ai-bill-of-rights/>.

³⁴ <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/>.

help the public discern between human and AI-generated content, which has raised controversy and criticism from deepfake detection researchers.”

In February 2024, the U.S. AI Safety Institute Consortium was established with 200 members from business and government. Their website, listed at the end of this paper, contains a resource center with comprehensive information about AI regulatory efforts.

- C. **Congressional Legislation:** One useful resource is the Brennan Center’s Artificial Intelligence Legislation Tracker³⁵, which includes brief descriptions of current bills and a direct link to their full text. The website states that the Center:

“...aims to increase public awareness of the myriad proposed regulatory approaches to AI legislation by serving as a repository of such AI-related bills introduced this session. Until now, information about AI legislation has been scattered across the internet or was accessible only through expensive legislative tracking services. Given both the known and unknown risks of AI, it is critical that the public have easy access to information on how lawmakers are attempting to address concerns. To maintain a reasonable scope, the tracker is limited to bills introduced during the 118th Congress that would do at least one of the following:

- Impose restrictions on AI that is deemed high risk
- Require purveyors of AI systems to conduct evaluations of the technology and its uses
- Impose transparency, notice, and labeling requirements
- Create or designate a regulatory authority to oversee AI
- Protect consumers through liability measures
- Direct the government to study AI to inform potential regulation.

The tracker also includes data protection bills that would significantly impact AI providers. The bills included in the tracker address some of the most serious risks posed by AI systems, such as perpetuating discrimination and bias, opaque and untested operating systems, giving inaccurate information, undermining privacy, and enabling disinformation and manipulation of images, video, and audio to influence elections.”

D. **Equal Employment Opportunity Commission (EEOC)**

The EEOC has been active in the field of AI. On May 18, 2023, the Commission issued a Technical Assistance Document titled “Select Issues: Assessing Adverse Impact in Software, Algorithms, and Artificial Intelligence Used in Employment Selection Procedures Under Title VII of the Civil Rights Act of 1964” (Title VII 29 CFR Part 1607).³⁶ While it does not have the force of law and is not binding, it nevertheless contains a helpful FAQ section stating among other things that employers who use algorithmic tools to screen applicants will be held to the same “adverse impact” standards (the “four-fifths rule”) as apply to human decision makers. This “rule,” which is actually a guideline, states that a selection rate for any race, sex, or ethnic group which is less than four-fifths (4/5) (or eighty percent) of the rate for the group with the highest rate will generally be regarded by the Federal enforcement agencies as evidence of adverse impact.

The Commission settled for \$365,000 its first AI case against a Chinese firm (iTutorGroup, Inc.) operating in U.S., which was using AI for screening applicants. According to the EEOC’s lawsuit, in 2020, iTutorGroup programmed their tutor application software to automatically reject female applicants aged 55 or older

³⁵ <https://www.brennancenter.org>.

³⁶ <https://www.eeoc.gov/laws/guidance/select-issues-assessing-adverse-impact-software-algorithms-and-artificial>.

and male applicants aged 60 or older. iTutorGroup rejected more than 200 qualified applicants based in the United States because of their age.³⁷

There is no current legal requirement to disclose to applicants that AI is being used in the background to evaluate their applications, nor any requirement that applicants consent to such use. The employer is ultimately liable for employment decisions and for any EEOC violations, regardless of the technology the employer uses.

E. California

1. *Executive order*: The Governor's September 6, 2023 Executive Order directs state agencies to adopt a proactive approach to AI by July 2024.³⁸ It states:

"To deploy GenAI ethically and responsibly throughout state government, protect and prepare for potential harms, and remain the world's AI leader, the Governor's executive order includes a number of provisions:

- Risk-Analysis Report: Direct state agencies and departments to perform a joint risk-analysis of potential threats to and vulnerabilities of California's critical energy infrastructure by the use of GenAI.
- Procurement Blueprint: To support a safe, ethical, and responsible innovation ecosystem inside state government, agencies will issue general guidelines for public sector procurement, uses, and required training for application of GenAI – building on the White House's Blueprint for an AI Bill of Rights and the National Institute for Science and Technology's AI Risk Management Framework. State agencies and departments will consider procurement and enterprise use opportunities where GenAI can improve the efficiency, effectiveness, accessibility, and equity of government operations.
- Beneficial Uses of GenAI Report: Direct state agencies and departments to develop a report examining the most significant and beneficial uses of GenAI in the state. The report will also explain the potential harms and risks for communities, government, and state government workers.
- Deployment and Analysis Framework: Develop guidelines for agencies and departments to analyze the impact that adopting GenAI tools may have on vulnerable communities. The state will establish the infrastructure needed to conduct pilots of GenAI projects, including California Department of Technology approved environments or "sandboxes" to test such projects.
- State Employee Training: To support California's state government workforce and prepare for the next generation of skills needed to thrive in the GenAI economy, agencies will provide trainings for state government workers to use state-approved

³⁷ <https://www.eeoc.gov/newsroom/eeoc-sues-itutorgroup-age-discrimination>.

³⁸ <https://www.gov.ca.gov/wp-content/uploads/2023/09/AI-EO-No.12--GGN-Signed.pdf>.

GenAI to achieve equitable outcomes, and will establish criteria to evaluate the impact of GenAI to the state government workforce.

- GenAI Partnership and Symposium: Establish a formal partnership with the University of California, Berkeley and Stanford University to consider and evaluate the impacts of GenAI on California and what efforts the state should undertake to advance its leadership in this industry. The state and the institutions will develop and host a joint summit in 2024 to engage in meaningful discussions about the impacts of GenAI on California and its workforce.
- Legislative Engagement: Engage with Legislative partners and key stakeholders in a formal process to develop policy recommendations for responsible use of AI, including any guidelines, criteria, reports, and/or training.
- Evaluate Impacts of AI on an Ongoing Basis: Periodically evaluate for potential impact of GenAI on regulatory issues under the respective agency, department, or board’s authority and recommend necessary updates as a result of this evolving technology.”

In response to this Executive Order, the California Department of Technology created an “Artificial Intelligence Community” (AIC), which meets quarterly. Deliverables include evaluation of generative AI risks and benefits, use in call centers, roadway safety for vulnerable users, traffic mobility, language access, and health facility inspections. The state recently issued “GenAI Guidelines for Public Sector Uses, Procurement, and Training”.³⁹ While the Guidelines apply only to state agencies, they can inform policies that cities may consider adopting.

2. *Legislature:* This section describes the bills introduced by California legislators as of the writing of this paper. The text is taken directly from the legislative record, edited for clarity.
 - a. AB 302 (1-26-2023, Ward) would require the Department of Technology, in coordination with other interagency bodies, to conduct, on or before September 1, 2024, a comprehensive inventory of all high-risk automated decision systems that have been proposed for use, development, or procurement by, or are being used, developed, or procured by, state agencies.
 - b. AB331 (1-30-2023, Bauer-Kahan) This bill died in committee, but is included for completeness and because it may resurface. It would have, among other things, required a deployer and a developer of an automated decision tool, both as defined, to annually perform an impact assessment for any such tool starting on January 1, 2025; notify any natural person that is the subject of the consequential decision that an automated decision tool is being used to make, or be a controlling factor in making, the consequential decision; provide that person with a statement of the purpose of the automated decision tool; accommodate a natural person’s

³⁹ [State of California GenAI Guidelines for Public Sector Procurement, Uses and Training.](#)

request to not be subject to the automated decision tool and to be subject to an alternative selection process or accommodation. This bill would prohibit a deployer, including local government agencies, from using an automated decision tool that results in algorithmic discrimination.

- c. SB 313 (2-6-2023, Dodd) This bill would enact the California AI-ware Act, which would establish, within the Department of Technology, the Office of Artificial Intelligence, and would grant the office the power and authority necessary to guide the design, use, and deployment of automated systems by a state agency to ensure that all AI systems are designed and deployed in a manner consistent with state and federal laws and regulations regarding privacy and civil liberties and that minimizes bias and promotes equitable outcomes for all Californians. This bill would require any state agency that utilizes generative artificial intelligence to directly communicate with a natural person to provide notice to that person that the interaction with the state agency is being communicated through artificial intelligence, and require the state agency to provide instructions to inform the natural person how they can directly communicate with a natural person from the state agency.
 - d. SB 721 (3-22-2023, Becker) This bill would, until January 1, 2030, create the California Interagency AI Working Group to deliver a report to the Legislature regarding artificial intelligence. The bill would require the working group members to be Californians with expertise in at least two of certain areas, including computer science, artificial intelligence, and data privacy. The bill would require the report to the Legislature to include, among other things, a recommendation of a definition of artificial intelligence as it pertains to its use in technology for use in legislation.
 - e. SB 896 (1-3-2024, Dodd) This bill, the Artificial Intelligence Accountability Act, would require the Government Operations Agency, the Department of Technology, and the Office of Data and Innovation to produce a State of California Benefits and Risk of Generative Artificial Intelligence Report that includes an examination of the most significant, potentially beneficial uses for deployment of generative artificial intelligence tools by the state. The bill would require a joint risk analysis of potential threats posed by the use of generative artificial intelligence to California's critical energy infrastructure, including those that could lead to mass casualty events and environmental emergencies. This bill would also require a state agency or department that utilizes generative artificial intelligence to directly communicate with a person, either through an online interface or telephonically, to clearly and in a conspicuous manner identify to that person that the person's interaction with the state agency or department is being communicated through artificial intelligence. This bill would also require an automated decision-making system utilized by a state agency or department to be evaluated for risk potential before adoption, as specified.
3. *Civil Rights Division:* The latest iteration of the CRD's proposed regulations⁴⁰ was released on February 10, 2023. The revised proposed regulations are designed to address decisions made by an AI (defined as machine learning systems that can make predictions or decisions) that has an adverse impact on applicants or employees based on a protected characteristic. The definitions also broaden the definition of an "agent" to an individual who uses an AI to make hiring or employment decisions. The

⁴⁰ Proposed Modifications to Employment Regulations Regarding Automated-Decision Systems.

regulations also include data retention requirements for AI systems used in making such decisions. As noted earlier, it will be imperative that the city have unfettered access to such data to defend itself from claims that the AI engaged in behavior that discriminated against applicants or employees.

4. *Local Policies:* Some California cities have adopted AI policies that address use by the entity and employees. San Jose has created a “GovAI Coalition” that includes members from across the country. The website contains resources such as template policies, manuals, vendor contracts, and a vendor registry, among many other useful features.⁴¹ San Francisco adopted guidance for city staff when using AI tools.⁴²

VII. Other Jurisdictions

- A. **New York:** New York was one of the first states to adopt a law regulating AI.⁴³ The New York City Bias Audit Law (Local Law 144) was enacted by the NYC Council in November 2021. The law went into effect on January 1, 2023, and enforcement was authorized to begin on July 5, 2023. Companies that are operating and hiring in New York City are prohibited from using automated tools to hire candidates or promote employees, unless the tools have been independently audited for bias.
- B. **Seattle, Washington:** On November 3, 2023, Seattle adopted a Generative AI Policy, described on its official city website⁴⁴ and excerpted as follows:

“The seven governing principles are: Innovation and Sustainability; Transparency and Accountability; Validity and Reliability; Bias and Harm Reduction and Fairness; Privacy Enhancing; Explainability and Interpretability; Security and Resiliency.

“The City’s new AI policy touches on many aspects of generative AI. It highlights several key factors to responsible use in a municipality, including attributing AI-generated work, having an employee review all AI work before going live, and limiting the use of personal information to help build the materials AI uses to develop its product. The policy also stipulates any work with a third-party vendor or tool must also include these principles for AI.

“This will help [mitigate] novel risks that have the potential to adversely affect the City’s ability to fulfill its legal commitments and obligations about how we use and manage data and information. City employees using AI technology will be held accountable for compliance with these commitments.

“All use of AI technology must go through the same technology reviews as any other new technologies. Those reviews take an in-depth look at privacy, compliance, and security, among others. ...

“The City policy applies to generative AI, which is a special type of AI technology. Generative AI produces new content for user requests and prompts by learning from large amounts of data called a ‘large language model.’ The capability to create new content, and to continually learn from these large data models makes it possible for a computerized system to produce content that looks and sounds like it was done by a human.

“While AI, including generative AI, has the potential to enhance human work across many fields of human enterprise, its use has also raised many questions about the consequences of employing

⁴¹ [Government AI Coalition | City of San José \(sanjoseca.gov\)](https://www.sanjoseca.gov/government-ai-coalition).

⁴² <https://www.sf.gov/information/guidance-city-staff-using-generative-ai-tools>.

⁴³ <https://www.nyc.gov/assets/dca/downloads/pdf/about/DCWP-AEDT-FAQ.pdf>.

⁴⁴ <https://seattle.gov/documents/Departments/SeattleIT/City-of-Seattle-Generative-Artificial-Intelligence-Policy.pdf>.

smart systems. Among these are ethics, safety, accuracy, bias, and attribution for human work used to inform AI system models.”

VIII. General Resources

These include academic studies and research across all disciplines, extensive glossaries, risk management frameworks, free newsletters, and invitations to webinars and seminars.

- A. Stanford Human-Centered AI⁴⁵
- B. Berkeley Center for Law and Technology⁴⁶
- C. Markkula Center for Applied Ethics⁴⁷
- D. National Institute of Standards, Technology; AI Resource Center⁴⁸
- E. Institute for Local Government (ILG) Webinar on AI⁴⁹
- F. Prior CalCities City Attorney papers:
 - 1. September 22, 2023 by Matthew C. Slentz, “Artificial Intelligence and the Practice of Law,” dealing with the risks of generative AI and use of AI tools by attorneys.
 - 2. September 22, 2023 by Peter Lee, UC Davis School of Law, “Promises and Perils of AI.”

⁴⁵ <https://hai.stanford.edu>

⁴⁶ <https://www.law.berkeley.edu/research/bclt/>

⁴⁷ <https://www.scu.edu/ethics/>

⁴⁸ <https://www.nist.gov/artificial-intelligence#:~:text=NIST%20contributes%20to%20the%20research,improve%20our%20quality%20of%20life>

⁴⁹ <https://www.ca-ilg.org/post/leading-local-artificial-intelligence-local-government>