Lozano Smith
ATTORNEYS AT LAW

VOLUME 2

RISKS OF AI

NAVIGATING THE LEGAL LANDSCAPE OF ARTIFICIAL INTELLIGENCE

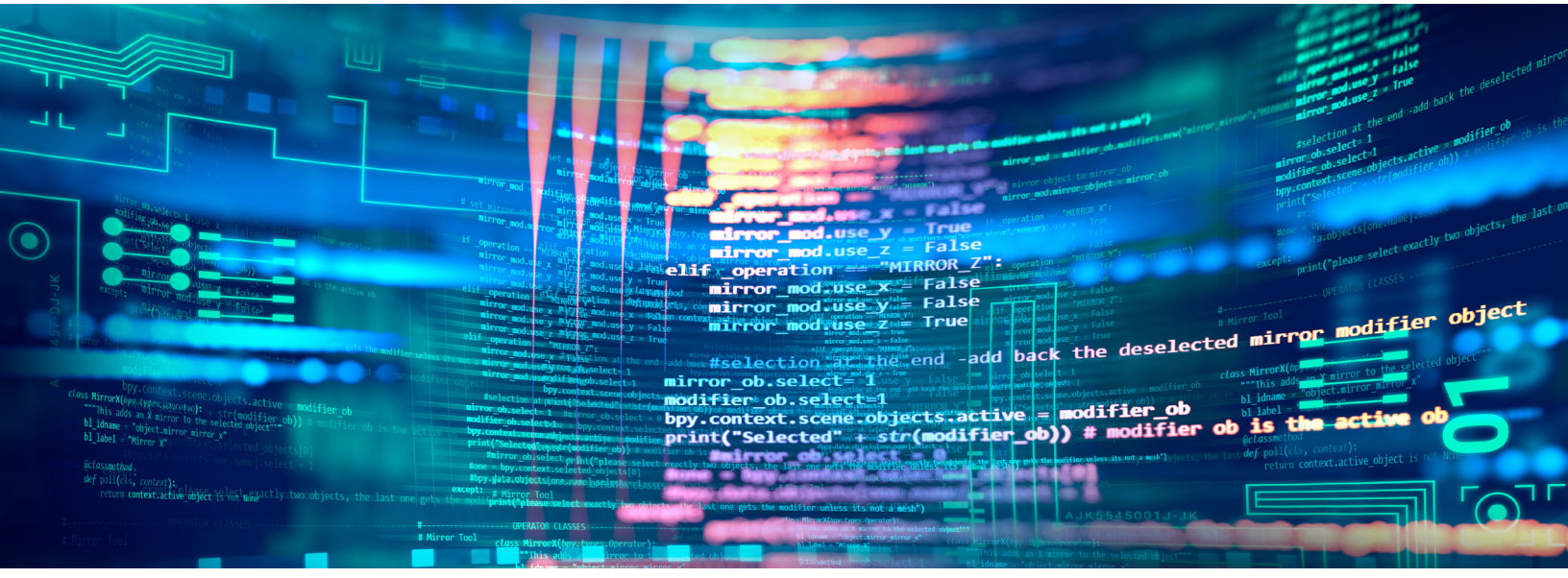
01 GENERAL RISKS OF AI

01A. BIAS

A primary concern raised about the use of AI systems is its potential to exhibit bias in decision making. AI bias, a reflection of discriminatory patterns or unequal representations in the data or its application, poses significant risks. It can inadvertently perpetuate existing societal prejudices, leading to unequal treatment of individuals or groups. Recent experiences with Google Gemini have also demonstrated that programming protocols intended to prevent bias, can also lead to the generation of inaccurate data.

AI is fundamentally shaped by the data it is trained on. This data, typically vast in scope, is generally all human generated and it does not always provide a neutral or unbiased representation of reality. AI systems learn by identifying patterns and correlations in their training data. If this data includes historical prejudices or societal biases – whether in terms of race, gender, socioeconomic status, or other characteristics – the AI is likely to inadvertently learn and replicate these biases. For instance, if an AI system is trained on employment data that historically favors a particular demographic, it may continue to replicate this favoritism, despite changes in societal norms or legal standards. The repercussions of training AI systems on biased data are significant, especially for public agencies. Decisions based on such data can lead to discriminatory outcomes, such as unfair resource allocation, biased hiring practices, or unequal service provision.

Not all AI bias is caused by the underlying biases that are learned by the model's training data. Bias can also occur due to the way the AI algorithm processes and prioritizes different inputs, sometimes at the direction of the user. In many AI systems, decisions are made based on a set of features or attributes considered relevant. For instance, in employment decisions, factors such as years of experience, education level, or past job



titles might be used. While these features appear neutral, they can inadvertently disadvantage certain groups. For example, prioritizing 'years of experience' heavily could lead to decisions that unfairly disfavor younger applicants, who may be equally skilled but have fewer years in the workforce.

The algorithms driving AI systems often assign different weights to various inputs, influencing the outcome. This prioritization, while designed to optimize decision-making, can unintentionally marginalize certain groups. For example, consider an AI system used by a public agency for allocating community development funds. If the algorithm prioritizes factors such as historical tax revenue or past project success rates, it may inadvertently disadvantage lower-income or historically underfunded communities. These areas, despite needing more resources, might receive less funding because the algorithm overlooks their potential for improvement and focuses on past performance metrics.

Proxy discrimination in AI occurs when an algorithm uses variables that, while not explicitly related to protected characteristics like race or gender, serve as stand-ins or proxies for these characteristics. This indirect form of discrimination can be particularly insidious because it often goes unnoticed, yet it can have profound impacts on fairness and equity. Proxy variables are attributes or factors that are not inherently discriminatory but are closely correlated with protected characteristics. For example, an AI system in a public agency might use zip code as a factor in decision-making processes, such as allocating resources or prioritizing service requests. However, since zip codes can closely correlate with racial and socioeconomic demographics, relying heavily on this factor could lead to decisions that inadvertently favor or disfavor certain groups based on where they live.

AI-DRIVEN DISCRIMINATORY FEEDBACK LOOPS OCCUR WHEN AI SYSTEMS, THROUGH THEIR DECISIONS AND ACTIONS, INADVERTENTLY REINFORCE AND AMPLIFY EXISTING BIASES OR INEQUALITIES.

AI-driven discriminatory feedback loops occur when AI systems, through their decisions and actions, inadvertently reinforce and amplify existing biases or inequalities. These feedback loops begin when an AI system makes decisions based on biased data or criteria. The outcomes of these decisions then become part of the new data set, which the AI continues to learn from, thereby reinforcing the initial bias. Over time, this cycle

TESTING AI MODELS FOR BIAS INVOLVES ANALYZING HOW THE SYSTEM MAKES DECISIONS ACROSS DIFFERENT GROUPS AND SCENARIOS.

can deepen existing inequalities or create new forms of discrimination. Imagine an AI system designed to identify students needing additional academic support. If this system is trained on data that inadvertently prioritizes certain indicators of performance – which may be influenced by socio-economic status, access to resources, or other external factors – it might consistently recommend more advanced resources for students from more affluent backgrounds while relegating those from underprivileged backgrounds to an academic intervention program. Over time, this can widen the educational achievement gap, as the AI's decisions reinforce and exacerbate existing disparities.

The nature of AI, particularly in advanced and complex systems, often involves a level of opacity that makes it challenging to understand how decisions are made, thus obscuring whether and how biased reasoning might be influencing AI-driven decisions. Extremely complex computations make it difficult to trace how inputs are transformed into outputs. In simpler terms, these systems can become 'black boxes' where the reasoning behind a specific decision is not transparent.

For public agencies, the inability to fully understand or explain the decision-making process of AI systems poses significant challenges. It raises questions about accountability and trust, particularly in scenarios where decisions have substantial impacts on individuals or communities. Without transparency, it is difficult to ascertain whether decisions are fair, free from bias, or even aligned with the agency's goals and legal obligations. The lack of transparency exacerbates the issue of identifying and addressing bias in AI systems. If the decision-making process is unclear, it becomes challenging to determine whether a biased reasoning pattern exists and, if so, what is causing it. This is particularly problematic in situations where decisions may be influenced by subtle forms of bias that are not immediately apparent.

For public agencies employing AI systems, it is vital to be aware of the potential for bias within these models and understand the general approaches for mitigating it. One strategy is early testing of AI models to identify potential biases. Testing AI models for bias involves analyzing how the system makes decisions across different groups and scenarios. This process helps identify if the AI system is unfairly favoring or disadvantaging certain groups. Since most public agencies may not have the technical capacity to test and correct AI models internally, it is advisable to engage with developers, vendors, or third-party auditors who have the necessary expertise. Additionally, agencies could require that entities they contract with conduct thorough bias testing as part of their service agreement. This can include periodic reviews and audits of the AI systems to identify and address any emerging biases. Public agencies can establish standards for the models they utilize, including requirements for transparency regarding bias testing procedures and corrective measures.

INCORPORATING A "HUMAN IN THE LOOP", REQUIRING HUMAN OVERSIGHT OF AI-GENERATED DECISIONS, IS A CRITICAL MITIGATION MEASURE FOR ADDRESSING POTENTIAL BIASES IN AI SYSTEMS, ESPECIALLY WITHIN PUBLIC AGENCIES.

When bias is identified after deployment, public agencies should first and foremost ensure that the bias does not impact agency decisions. Additionally, agencies can work with developers to adjust the model or seek alternative solutions that demonstrate a stronger commitment to unbiased outcomes. While the technical details of these corrections are managed by specialists, agency officials should understand the broad strategies used for bias correction, such as diversifying training data or adjusting the algorithm's decision-making parameters.

Incorporating a "human in the loop," requiring human oversight of AI-generated decisions, is a critical mitigation measure for addressing potential biases in AI systems, especially within public agencies. This approach ensures that AI-generated decisions are not made in isolation but are instead reviewed and, if necessary, adjusted by human oversight. For this approach to be effective, agency personnel involved in overseeing AI decisions need appropriate training. They should be aware of the potential biases AI systems might hold and be equipped to identify signs of such biases in decision outputs. Additionally, fostering a culture of critical engagement with AI tools is essential, encouraging employees to question and challenge AI recommendations when necessary. Agencies should also have protocols for reporting potential biases or inaccuracies in AI decisions. This feedback can inform the agency's decision on whether to use a specific AI system and may also be communicated to the developer to enhance the model, aiming to mitigate bias.

Bias may also be minimized by engaging in prompting strategies which promote transparency and are less prone to bias. Instead of asking an AI system to make broad assessments or decisions, a more targeted approach in questioning can

"THE THING THAT'S GOING TO MAKE ARTIFICIAL INTELLIGENCE SO POWERFUL IS ITS ABILITY TO LEARN, AND THE WAY AI LEARNS IS TO LOOK AT HUMAN CULTURE."

- DAN BROWN

reveal the underlying logic of its conclusions. For example, rather than having an AI model determine which students should be placed in an academic intervention program, a more effective prompt would be to ask the AI model to identify students struggling with specific aspects of the curriculum, such as multiplication, and to explain the factors the AI system considered to reach that conclusion. This process could be repeated over various aspects of curriculum, and at that point the AI model could be asked to identify students in need of an academic intervention program and to explain the factors the AI considered for each student. This approach not only provides specific insights but also enables human reviewers to understand the basis of the AI's decisions. While the AI system may not always be transparent in explaining its decisions, this approach allows the human reviewer to assess whether the AI system's decisions consider relevant factors or mask other influences. Prompting strategies that take a targeted approach to decision making, and require AI systems to explain their logic, empower human decision makers to make informed decisions about whether to accept, modify, or reject AI recommendations. It also fosters a more critical and engaged approach to using AI tools, ensuring that these powerful technologies are used responsibly and ethically.

In navigating the complexities of AI, public agencies should proactively address the inherent risks of bias through rigorous testing, human oversight, and strategic prompting. By implementing these measures, agencies can harness the potential of AI while upholding their commitment to fairness, transparency, and accountability in public service.

01 B. PRIVACY

Like any software application or cloud service, users must be able to trust and rely on the companies they are entrusting with sensitive information and ensure there are adequate data protections. This concern is particularly acute in the context of AI and LLMs, as user inputs and responses may be used to train current or future models. One of the ways that AI models advance is through reinforcement learning, discussed in more detail above, whereby the model generates one or several responses and a human user rates the output. The model incorporates human feedback and adjusts the model to perform closer to the desired result in the future. User interactions with LLMs and other AI systems can be used to provide some of this reinforcement learning, as users can either directly provide feedback on a model's response to a prompt (e.g., a "thumbs up" or "thumbs down" in the ChatGPT interface) or the AI system administrator may be able to discern whether the user was satisfied with the response based on subsequent interactions (e.g., the user thanks the model for the help, or has to repeatedly clarify their questions and prompt to get to a desired output). Real world users providing this kind of feedback is very valuable for future AI development.

As discussed in Volume 1, LLMs typically do not store information in a database or memorize the data they're trained on. Instead, they learn from patterns within the data. So, while a model trained on user inputs shouldn't directly reproduce those prompts for other users, it can still understand and replicate the underlying information. In other words, it may generate responses similar to those

REAL WORLD USERS PROVIDING AI SYSTEMS WITH FEEDBACK IS VERY VALUABLE FOR FUTURE AI DEVELOPMENT.

it was trained on when faced with similar queries from different users.

Imagine a public agency implementing an AI-powered chatbot to assist citizens with inquiries. An employee engages with the chatbot to input diverse citizen questions and receive appropriate responses. Although the chatbot does not retain the exact questions or answers, if the employee identifies a unique citizen concern and seeks clarification, the chatbot may recognize similar issues in subsequent interactions and offer tailored assistance.

While LLMs are generally trained to learn the patterns of the data they are trained on, rather than to "memorize" exact copies, there are instances when this can happen. LLMs memorizing their training data is called "overfitting" and is a problem developers guard against as it makes the models less useful. However, in rare circumstances the models may memorize portions of their training data and reproduce it in an output to a user request. One study found that an attacker could cause ChatGPT and other LLMs to reproduce verbatim training data in an output, including personally identifiable information. The New York Times used ChatGPT outputs which included verbatim text from New York Times articles as part of the basis of their copyright lawsuit against OpenAI and Microsoft. While the risk of information contained in user inputs being outputted to other

users is low, there is still some risk of exposing sensitive information. Accordingly, public agencies should avoid including confidential information, particularly personally identifiable information, in LLM inputs.

By default, most publicly available models utilize user interactions with LLMs for training, potentially incorporating user prompts without explicit consent for training purposes. Some LLM providers provide the ability to "opt-out" of prompts being used for training purposes or offer specific enterprise grade plans with higher security guarantees, including that user data is never used for training purposes.

Even if the agency or individual user has "opted out" of having their data used for training, the same privacy concerns that an agency would consider with any other software or cloud service would still apply. Agencies should still evaluate LLM services based on their security practices, compliance with applicable data privacy laws, and ability to prevent and respond to data breaches.

LLMs which run exclusively on the user or agency's own hardware, generally do not have these same kinds of privacy concerns. As discussed in Volume 1, there is an increasing variety of LLM models that can run entirely on an agency's hardware, or even individual user's personal computers. As all of the computing and inference is done on the local hardware, no user prompts or data needs to be transmitted to the developer's servers. Thus, the developer does not have the opportunity to train on the user data nor do they store user data which could be vulnerable to a data breach.

Public agencies should ensure the AI systems they utilize provide sufficient data-protection, such as refraining from training on user prompts, prior to inputting sensitive information. This precaution extends to software incorporating AI features, as user data might be directed to third-party AI models for processing and response. If the agency cannot be sure of the data-privacy protections of the AI system, sensitive and confidential information should not be entered into the models, particularly personally identifiable information (names, addresses, identifiers, etc.). This is especially true for school districts related to student data.

EXAMPLE

Utilizing AI to draft individualized education program ("IEP") goals may result in a Family Educational Rights and Privacy Act ("FERPA") violation if a student's personally identifiable information is disclosed to an AI system.

01 C. ACCOUNTABILITY

While AI can significantly enhance efficiency, decision-making, and service delivery, it also brings forth critical questions of accountability. The common theme of this section is that a public agency is ultimately responsible for their decisions, even if AI systems are utilized to inform or advise on decision making. Accordingly, it is critical for humans to be informed about the limitations and risks of AI, oversee all uses of AI and independently evaluate outputs, and ultimately come to independent decisions informed by, but not reliant on, information and recommendations produced by AI. This is not only critical to ensure agency decisions are well reasoned and legally permissible, but it provides the foundation of transparency that is imperative for ensuring public trust in AI deployments.

The cornerstone of integrating AI into public agency operations is the recognition that, despite the advanced capabilities of these technologies, human decision-makers hold the ultimate responsibility for outcomes. This principle is pivotal in ensuring that AI serves as a tool for enhancement, not as a replacement for human judgment and accountability. As discussed in this section and elsewhere in this series, while AI systems are powerful and likely to provide a myriad of benefits to public agencies, they are still prone to errors, bias, and lack human wisdom, understanding, and ethical reasoning. Agency staff utilizing AI systems must maintain a "human in the loop" actively evaluating AI outputs to identify and prevent errors. Staff and decision makers alike need to be informed about the benefits and limitations of AI systems and should only ever use AI outputs as

a single data-point in a comprehensive decision-making process. Unquestioning reliance on AI outputs may present a variety of legal risks for public agencies.

Accountability is crucial from the outset, encompassing decisions on AI implementation, its intended use, and the selection of AI systems. These initial steps set the foundation for how effectively a public agency can harness the power of AI while maintaining the trust and confidence of the public it serves. Organizational accountability is essential and includes decisions on the deployment, purpose, and choice of AI systems utilized by a public agency. These early decisions lay the groundwork for effective use of AI and it is critical for agencies to develop clear policies on AI adoption and usage within their operations. Without these guidelines, there is a risk of employees independently using AI for various tasks, potentially exposing the agency to risks such as privacy breaches, bias in decision-making, and regulatory non-compliance. When deciding to implement AI systems, it is crucial to establish specific protocols that define who can use AI systems and for what purposes. Moreover,

offering thorough training to employees is essential to prevent accidental misuse and any liability associated with that use.

It is also important for agencies to be informed about the variety of AI systems to select the best system for the agency's needs. For example, open-source models deployed on an agency's own servers may provide certain data privacy benefits and be more appropriate to the extent the agency expects these systems to work with sensitive information, though currently available open-source models perform at a lower level than the major commercial models. It depends on the service, however, most paid AI services allow for users to opt out of having their prompts used for training purposes. Monitoring and controlling employee use of AI is possible when the employee is using employer provided computer equipment or accounts; otherwise it would prove difficult and an employer would need to rely on any employee policy that is in place. If the model will be deployed in a public or student facing way, it is important to consider what protections are available to ensure the model does not produce inappropriate, offensive, or harmful outputs. Does the agency have mechanisms in place to alert them of any inappropriate use of the AI or to notify them if the model generates a prohibited output? Is the agency able to test the models for potential biases and are there mechanisms to correct such biases? These are just some of the considerations the agency should be engaging in when evaluating AI deployments.

After models are deployed, it is critical that the agency continually evaluate the model's performance for accuracy and effectiveness. The agency may find that their LLM deployment excels in certain tasks, saving the agency staff time and resources, while it provides poor or inaccurate results in other tasks. The agency may find that certain prompting strategies produce better

AS AI BECOMES THE NEW INFRASTRUCTURE, FLOWING INVISIBLY THROUGH OUR DAILY LIVES LIKE THE WATER IN OUR FAUCETS, WE MUST UNDERSTAND ITS SHORT- AND LONG-TERM EFFECTS AND KNOW THAT IT IS SAFE FOR ALL TO USE.

- KATE CRAWFORD

results, whereas some tasks are just too complex for the current AI system to provide effective assistance. Continually monitoring and evaluating AI performance ensures that the agency can adapt their use of AI systems (or evaluate whether other systems are more appropriate). In addition to the general liability risks discussed in this document, an agency which deploys but fails to appropriately monitor an AI system could be found liable under a negligence theory. This liability may arise if the AI model generates harmful or inappropriate responses, whether directed towards a member of the public, a student, or an employee relying on the AI system for decision-making. Such liability could be attributed to the agency's failure to oversee the deployment adequately, especially if the potential risks were foreseeable with proper monitoring.

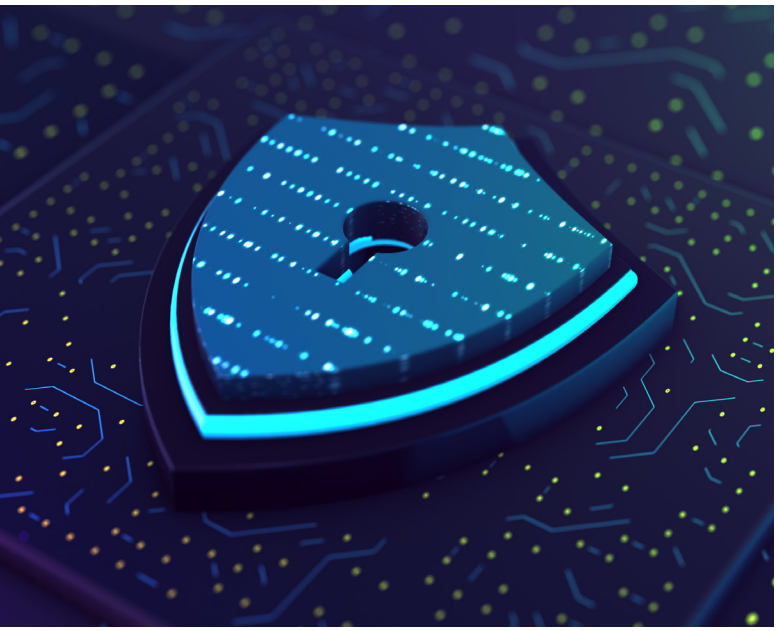
Related issues include overreliance and the subsequent degradation of human skills. Overreliance refers to situations where AI systems are excessively depended upon for decision-making, potentially leading to diminished human engagement and a failure to critically assess AI outputs. This reliance can result in a lack of questioning of AI-generated results, allowing unchallenged errors or biases to influence decisions. Moreover, an important concern that arises with the overuse of AI is the risk of skill degradation among staff. When employees rely heavily on AI tools, they may experience a decline in critical skills, particularly in areas requiring complex analysis and decision-making. This degradation not only diminishes individual capabilities but also impacts the overall resilience and adaptability of the agency. These issues are very related, as overreliance can lead to skill degradation, and skill degradation can lead to overreliance. In addition to this situation leading to poor decision making and performance of the public agency, it may also lead to legal liability to the extent that the agency implements flawed, biased, or legally non-compliant decisions based on AI advice. It is thus critical for agencies to implement protocols to guard against these risks.

WHEN EMPLOYEES RELY HEAVILY ON AI TOOLS, THEY MAY EXPERIENCE A DECLINE IN CRITICAL SKILLS, PARTICULARLY IN AREAS REQUIRING COMPLEX ANALYSIS AND DECISION-MAKING.

AI system stability and long term sustainability is also important to consider, particularly in the context of overreliance and skill degradation. While still in the infancy of widespread AI development and deployment, there are a large number of companies attempting to build and develop businesses and AI products, some of which may not remain viable over time. Over the Thanksgiving break in 2023, OpenAI saw its CEO fired and the vast majority of its staff threatened to quit in protest, creating concern among the business built upon or simply relying on OpenAI's services. While that situation was ultimately resolved without interruption of OpenAI's services, the episode illustrates that agencies need to think critically regarding the extent they intend to rely on any one AI service and have backup plans in the event a service is no longer available. Boardroom conflicts are just one of the potential events that could lead to the discontinuation of AI services. AI models are trained on vast amounts of information primarily

obtained on the Internet, a significant percentage of which is copyrighted. A number of high-profile lawsuits have been brought challenging these companies' use of this copyrighted work. If these lawsuits are successful, it is possible that these AI services may suddenly be unavailable, causing hardship for agencies that have come to rely on them. Locally deployed open-source models have the advantage of being immune from issues with the viability of the creators of those models, but if the models themselves are ruled to contain "copies" of protected work, an agency's deployment of such models may subject the agency itself to potential liability.

Another crucial issue is public transparency regarding AI use. As the public becomes more aware of and familiar with AI systems, and the related risks, it is likely there will be increased scrutiny on how agencies are (and are not) utilizing AI in their provision of services. Agencies should be proactive in making the public, constituents, customers, and parents and students aware of how the agency is utilizing AI and what protocols and safeguards are in place. Doing so can help alleviate skepticism surrounding the agency's use of AI and mitigate the perception that the agency is concealing its AI practices, which could lead to negative consequences. In addition to general transparency on AI deployments, agencies should be especially careful and transparent about public and student facing AI deployments given the risk of harmful or inappropriate AI outputs. For example, if a city deploys an AI chatbot to assist users in identifying applicable building and municipal code requirements but does not prominently warn the users that the AI may produce inaccurate responses, and the AI model produces an inaccurate output which is relied upon by the user to his or her detriment, the user may have a legal action against the agency.



Similarly, use of AI systems by public agency officials and employees may result in the creation of "records" for purposes of the California Public Records Act. This is very much an untested area of the law, and whether a particular interaction with an AI system results in the creation of a disclosable "record" under the California Public Records Act may depend on a number of factors such as how the system is deployed (locally run open-source model vs commercially available model which uses user prompts for training) and how the model was used. Agencies should still take into consideration the potential need to disclose both user prompts and AI outputs in response to Public Records Act requests. This underscores the central importance of establishing protocols on AI use, educating staff and officials on permitted uses and risks and limitations of AI

systems, and being transparent with the public regarding AI deployments as early as possible.

Agencies should also be aware that the regulatory landscape of AI development and deployment is changing rapidly. Federally, many departments and agencies, including the Department of Education, are set to put out regulations on the use of AI systems at some point in 2024. The State of California is similarly considering various regulatory and statutory measures that could affect the use of AI in public agencies. Accordingly, it is critical for agencies to stay informed about these developments to avoid any violations of law or regulation in this quickly evolving legal landscape.

EXAMPLE

An employer will be liable if an algorithmic decision-making tool utilized in the hiring process results in a violation of a provision of the Americans with Disabilities Act.

01 D. IMPACTS OF WIDESPREAD AI ADOPTION

It is important to recognize that the impacts of AI on public agencies will not just be limited to how agencies themselves chose to deploy (or not) these systems. As AI begins to proliferate through society, public agencies will have to learn to adapt to new challenges these systems bring.

As constituents and parents increasingly turn to AI systems for information regarding agencies, policies, and laws, as well as to shape their understanding of the agency's decisions, reliance on these systems is expected to rise. Although the information people receive from LLMs may be incorrect for any number of reasons (bad prompting, hallucinations, poor quality information retrieval), people are primed to believe that these systems are providing accurate information, particularly when it comes from companies they have high confidence in (e.g., Google, Microsoft). For example, many people have come to trust the accuracy of information obtained from a Google Search result and may incorrectly assume the information obtained from a Google Gemini interaction is of equal quality. It is likely that agencies will begin to see misinformation spread in their communities based on shared interactions with LLMs, a fact that will be further exacerbated by people's biases for trusting AI sources.

AI can also be used to spread disinformation intentionally. This includes things like mass producing targeted social media posts, voice cloning systems being used to make it appear that an official said something they did not, and full image or video deepfakes of public agency officials. As we head into the election season, AI disinformation is one of the primary concerns of experts studying the impact of AI on society. Given AI's potential to make these disinformation campaigns easier, we can expect these strategies to be employed at the local level as well as state-wide and federal elections. This is particularly likely as political activists and groups have increasingly targeted their activism at the local level, a strategy that first gained prominence during the COVID-19 pandemic.

It is also possible that as the general public gains an understanding and acceptance of AI technology, their expectations about public agency use of these systems may shift. Currently, given the attention on hallucinations, bias, and other problems associated with AI systems, many communities may be skeptical of any public agency deployment of AI technology. However, it's possible that over time, as AI systems

improve and are integrated into the lives of more people, these perceptions may shift. Agencies may experience the general public sentiment to shift from "Why are you using AI given all of its flaws" to "Why aren't you using AI given its benefits, particularly in terms of cost savings." For example, the public may question why agency resources and staff time are being used for a certain task when available AI systems could significantly decrease the time and cost required for completion.

AI will also impact day-to-day operations of agencies, as the vendors they deal with increasingly utilize these systems both in their product offerings and in their own internal operations. For example, a vendor of finance software may utilize AI in an effort to provide better financial planning strategies without directly informing the client. The agency could thus find itself faced with many of the problems associated with AI information and decision-making discussed above, but without the opportunity to evaluate the risks. As an example of how a vendor's internal use could impact

AGENCIES SHOULD EXPECT THEY WILL INCREASINGLY BE UNDER SCRUTINY BY INDIVIDUALS AND GROUPS LEVERAGING AI FOR THEIR OWN PURPOSES.

agencies, a vendor could use AI to sneak an unfavorable contract provision into an agreement in such a way that the agency might not notice it on a quick review.

There is also the potential for groups and individuals to leverage AI systems to assess legal vulnerabilities of public agencies. LLMs can be used to analyze vast amounts of information about public agencies, whether publicly available on the internet or produced in response to CPRA requests and compare it against databases of laws and agency policies to identify legal vulnerabilities. No agency is perfect and minor legal missteps are not uncommon. AI systems will be able to thoroughly scan for these issues and identify concerns that likely would have been overlooked by a non-expert human. The purpose of this use could be fundamentally fair, such as a disappointed bidder using AI to scan the low bid for an agency contract in order to identify grounds to submit a bid protest. Another relatively innocuous example would be a non-profit concerned with compliance in certain areas of law (e.g., prevailing wage regulations) using AI to identify potential instances of agency non-compliance. However, this practice could be "weaponized" by individuals or groups looking to target a particular agency. Agencies should

expect they will increasingly be under scrutiny by individuals and groups leveraging AI for their own purposes.

Another likely risk category of AI proliferation is the increasing danger of cyber threats from bad actors. Current LLM systems are very capable of computer programing and could be utilized to develop malicious code used to target public agency data infrastructure. This means that even relatively unsophisticated bad actors would be able to deploy viruses, malware, ransomware, and other vectors of attack with relative ease. More importantly, the same machine learning principles that make LLMs and AI so effective, could be leveraged to increase the effectiveness of attacks. Imagine a virus that is able to learn from how its targets are able to neutralize prior attacks and autonomously adjust its code and methods to prevent that neutralization strategy in the future. This virus could also initially spend time on the victim's system probing for vulnerabilities and deploying an attack strategy specifically tailored to the target's unique system vulnerabilities. Just as AlphaGo was able to learn from its experience playing the game "Go" and to come up with unique gameplay strategies no human had yet thought of (discussed in Volume 1) so could a malicious AI threat vector deploy surprising attack methods.

Importantly, whereas corporations, public agencies, and the general public are largely engaging in a deliberative and thoughtful process around AI deployment, it is very likely that bad actors looking to leverage these new technologies are already actively developing new attack vectors relying on advances made in machine learning and AI generally. Accordingly, this may be an area of concern that comes to the forefront relatively rapidly. It is also important to note that this issue does not just affect public agencies and their systems directly but will also affect the vendors agencies rely on for various services. It is not only essential for public agencies to assess their own systems for potential vulnerabilities, but also to ensure the vendors they work with are aware of and preparing for these future threats so as to ensure agency operations are not compromised by vendor downtime and that agency data is not exposed through a breach of the vendor's systems.

Many of these challenges are not new. Agencies are familiar with unscrupulous vendors using one-sided and sometimes deceptive contract terms to gain advantage. Agencies are certainly familiar with CPRA requests and how they can sometimes be "weaponized." Cyber threats, particularly ransomware attacks in recent years, have been at the forefront of public agency data infrastructure challenges. What will be new is the speed and complexity of these issues moving into the future. Bad or even just troublesome actors will be able to utilize AI systems to supplement their own abilities, hypercharging the challenges agencies already face. Accordingly, even if an agency itself decides not to implement AI in its own operations, they will still have to adapt to a new AI centered reality.

AI IS A TOOL. THE CHOICE ABOUT HOW IT GETS DEPLOYED IS OURS.

-OREN ETZIONI

---

01 CONCLUSION

As the AI landscape evolves rapidly, the emergence of unforeseen harms and risks is inevitable. Recognizing this, both state and federal governments are moving towards implementing regulations to mitigate potential risks associated with AI systems. At Lozano Smith, we are dedicated to maintaining our position at the forefront of addressing legal issues related to AI. With our team of subject matter experts, we are committed to assisting your agency in navigating the complexities of AI law, ensuring that your policies, procedures, and utilization of AI systems adhere to relevant legal standards.

KEY CONTACTS

Karen M. Rezendes
Partner | Walnut Creek
krezendes@lozanosmith.com

Nicholas J. Clair
Senior Counsel | Sacramento
nclair@lozanosmith.com

Robert A. Lomeli
Associate | San Luis Obispo
rlomeli@lozanosmith.com

Karina Demirchyan
Associate | San Luis Obispo
kdemirchyan@lozanosmith.com

Andy Garcia
Executive Director | Fresno
agarcia@lozanosmith.com

ADDITIONAL RESOURCES

- Visit our website for the latest news and resources on the topic of artificial intelligence. > Click here
- Listen to our latest podcast on artificial intelligence. > Click here

Lozano Smith
ATTORNEYS AT LAW

LozanoSmith.com