

How Artificial Intelligence Combats Fraud and Cyberattacks



GOVLOOP
E-BOOK
2019

CLouDERA



Executive Summary

Government data is growing exponentially. Its volume, velocity and variety require advanced data analytic capabilities to empower informed decision-making, but many agencies are unequipped to manage such a large-scale task.

For an idea of how rapidly government datasets are expanding, Data.gov launched in 2009 hosting [just 47 datasets](#). Today, 10 years later, the site houses over 200,000 datasets from hundreds of sources. And Data.gov only represents the government data made available for public use.

Behind the scenes, federal agencies use metadata and data analysis to detect fraud, waste and abuse of federal programs and citizen services. Understanding and acting on these analytical insights derived from government data resources has the potential to support agency objectives in new and varied ways.

Additionally, fraudsters and cybercriminals are constantly outsmarting the overworked human personnel trying to thwart them. With the everyday bustle of modern government IT departments, it's difficult for teams to adapt to the changing tactics of nefarious entities attempting to take advantage of systematic weaknesses.

But the opportunity to use this data to combat fraud and cyberattacks requires sharing and analyzing data seamlessly across traditionally siloed departments and the manual monitoring of data-sharing, fraud detection and cyberattacks is time-consuming and expensive.

Fortunately, new technology is available to assist these efforts. Artificial intelligence (AI) and machine learning (ML) are the way forward as the amount of data continues to grow alongside the shifting threat landscape.

By analyzing the federal government's existing data landscape, AI and ML technologies can learn to detect system abnormalities for manual assessment. This frees up time for cybersecurity experts to focus on innovating new solutions instead of simply monitoring the perimeter.

By first integrating an agency's data onto a holistic, open-source platform with the ability to manage massive data infrastructures, the agency is then able to introduce an AI- or ML-based data analytics tool to detect adversaries. AI and ML are also scalable to adjust to the shifting security perimeters and new scams reshaping the threat landscape regularly.

In *How Artificial Intelligence Combats Fraud and Cyberattacks*, we explain how leading governments are working to integrate AI and ML into their analysis and technology to secure sensitive data and advance innovation.

Federal AI and Cybersecurity at a Glance

While AI and ML technology are still relatively new to the public sector, policies at the federal level have come a long way since the first research and development (R&D) initiatives. These stats and the accompanying timeline show the history of federal AI and ML policies and insights into federal workforce perceptions of the emerging technologies.

59%

of [federal employees](#) believe that intelligent technologies such as AI and ML will reduce repetitive tasks and administrative burden, 53% believe it will improve their productivity and 46% believe it will reduce errors.

80%

of [federal executives](#) said that within two-plus years, AI will work next to humans as a coworker, collaborator and trusted adviser.

“America’s national security and our global competitiveness depend on our ability to remain at the forefront of future technology advancements.

Investing heavily into AI is an investment in the future of our country.”

- [Pete Olson](#), Texas Rep.

The federal government’s [investment in unclassified R&D for AI and ML technologies](#) grew over 40% between 2015 and 2018.

\$688 million

[budgeted](#) for NIST to conduct AI research in fiscal year 2020.

\$208 million

[budgeted](#) to scale and develop DoD’s JAIC in fiscal year 2020.

\$4.9 billion

[budgeted](#) in unclassified AI and ML-related R&D spending for fiscal year 2020.

“Agencies should invest in R&D to increase the security and resilience of the Nation’s critical infrastructure from both physical threats and cyberattacks, which have increased rapidly in number and complexity in recent years.”

- [Mick Mulvaney and Michael Kratsios](#)

1.2 billion

hours of federal labor [could be saved](#) by implementing AI to automate government tasks.

\$41.1 billion

billion could be saved by [implementing AI](#) to automate government tasks.

October 2016

The National Science and Technology Council (NSTC) releases the first [National AI R&D Strategic Plan](#) under President Barack Obama as a framework for federally funded AI priorities. The plan was accompanied by a report, [Preparing for the Future of Artificial Intelligence](#), that surveyed the current state of AI, outlined possible applications and raised questions about possible implications.

December 2017

President Donald Trump's [National Security Strategy](#) becomes the first to call out the importance of AI for the future of the American military.



2018

The [National Defense Strategy](#) commits to investing in military applications of AI and ML, and the Department of Defense (DoD) releases its own [AI strategy](#) focusing on implementing AI to advance security.



May 2018

The White House charts a [Select Committee on AI](#) under NSTC to advise on AI priorities, consider private sector partnerships and identify opportunities.

June 2018

DoD announces the establishment of the [Joint Artificial Intelligence Center](#) (JAIC) to accelerate the delivery of agencywide AI services related to its security missions.



February 2019

The Trump administration releases the [American AI Initiative](#) and [Executive Order 13859: Maintaining American Leadership in AI](#) to encourage federal investment in AI to enhance security, among other initiatives. Meanwhile, the Intelligence Advanced Research Projects Activity (IARPA) announces two programs to promote use of AI in cybersecurity. [Secure, Assured, Intelligent Learning Systems](#) (SAILS) investigates ways to train AI systems to stop attackers from revealing personally identifiable information (PII) and [TrojAI](#) deploys software to scan algorithmic outputs for Trojan horse attacks.

June 2019

An [update](#) to the National AI R&D Strategic Plan includes a new focus on creating effective partnerships between the private sector, government and academia to generate technological breakthroughs.



July 2019

The National Institute of Standards and Technology (NIST) [releases a draft plan](#) for federal government engagement in advancing AI standards.



The How and Why of AI Fighting Fraud and Cyberattacks



As federal agencies continue toward hyper-connectedness and embrace digital transformation trends, enterprise risk continues to increase. From cyberthreats, to fraud, to ever-changing compliance regulations, today's government needs to increase visibility and unlock the power of ML and advanced analytics to lower enterprise risk.

Broadly, AI is the overarching term for any program that can complete a task based on existing data and algorithms. ML is a form of AI that has the ability to make decisions and constantly self-improve without specific programming adjustments. AI and its subset ML can both be used to automate labor and leave workplaces with more time to innovate solutions.

The prevalence of AI and ML in federal-level conversations makes right now an advantageous time to embrace the incoming wave. The Government Accountability Office (GAO) named AI one of [the top five emerging technologies](#) in 2018, and the current legislative push for modernization is an opportunity to harness it.

[The Federal Information Technology Acquisition Reform Act \(FITARA\)](#) scores federal agencies on a series of eight guidelines, including modernizing government technology, improving data efficiency and cybersecurity. FITARA is an opportunity for technologies such as AI and ML because it provides a chance to improve data management through emerging technologies that help combat cyberattacks and fraud.

Further, AI and ML embrace the [American AI Initiative](#) as set forth by a Trump administration executive order. This order encourages federal agencies to invest in AI R&D, notably to improve security, and creates committees dedicated to fleshing out AI strategies.

No technology can replace the complexity of human thought, but automation can leave agencies with more time for employees to live up to their full potential. AI and ML automation can be applied to tedious tasks that take staff away from more meaningful work. By limiting instances that need human attention, there's more time for activities such as developing strategic plans and mentoring colleagues.

In the cybersecurity sphere, AI and ML technologies are able to evaluate what the standard operating procedures of an agency look like to flag unusual actions for manual review. This relieves some of the pressure on existing cybersecurity personnel because they are able to strategize ways to overcome threats instead of carefully reviewing every action.

With this technology in place, valuable human labor hours will be freed up to work on more innovative projects and find creative ways to adapt to the evolving threat landscape. While personnel will never be completely out of the cybersecurity equation, their brainpower and innovative capabilities can be used more effectively if AI and ML techniques are put in place to detect fraud and cyberattacks at the federal level.

AI & ML Fight Fraud

Federal agencies manage and process heaps of data in their daily delivery of citizen services. From individual personally identifiable information to tracking the history of services provided, government personnel sort through massive amounts of information each day. Reviewing and cross-referencing this data to ensure that every constituent receives the correct help in a timely manner is an important part of the public sector's duties, but human error still occurs.

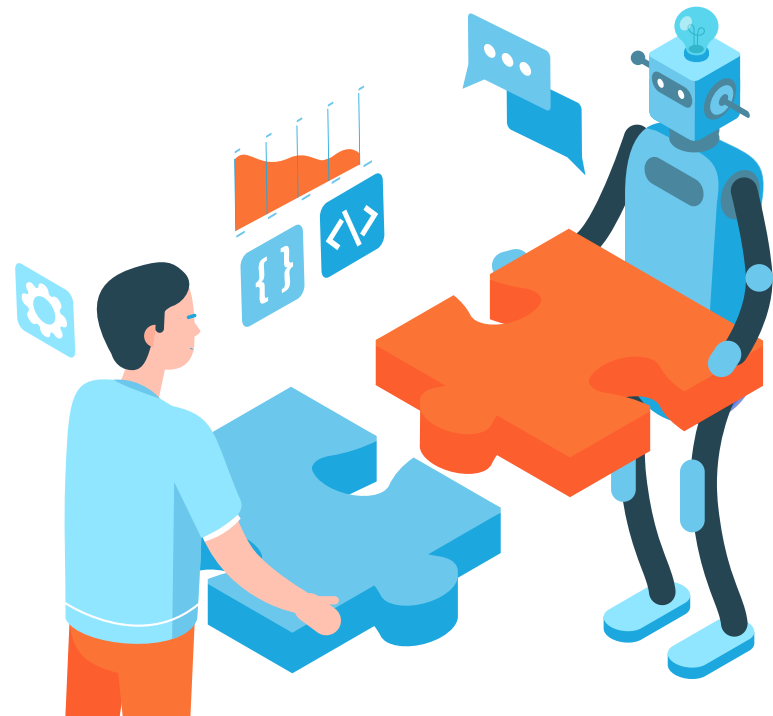
Consider, for example, small-scale fraud among government assistance programs, where ineligible citizens lie to qualify, or citizens misrepresenting their assets to avoid paying taxes. Employees attempting to fulfill their endless duties can miss signs of these attempts to throw off the balance of taxpayer funds and government commitments.

In 2018, the federal government lost over \$5 billion in confirmed fraudulent activities. This isn't just a concentrated attack on a single agency or program. The ability to take advantage of federal assistance spans across practically every agency, from the Department of Homeland Security (DHS) to the Veterans Affairs Department (VA).

In cases of fraud, data needs to be quickly examined and compared to other data before an ill-intentioned actor is able to take advantage of the system. Manual comparison of datasets to detect inaccuracies and imputations is possible, but it's time-consuming and a costly application of limited labor resources. With an automated fraud detection system in place, employees have more energy to focus on remediating the errors instead of simply detecting them.

The algorithmic capabilities of AI and ML can be programmed to detect fraudulent data based on previous patterns over time. After reviewing what years of honest interactions with the system look like, they can flag instances that seem out of the ordinary or deceptive. Then, staffers can manually review those use cases to make the final decision. Humans are still an important part of the process, but their competency is applied more efficiently.

From there, models can continue to improve. Based on what types of interactions the AI and ML systems flag and research from employees, the systems can be adapted to detect new fraudulent schemes within the evolving threat landscape. Instead of retraining the entire staff on how to catch new scams, the algorithm is adjusted to fit the new criteria.



AI & ML Fight Cyberattacks

Cyberattacks, including ransomware and data breaches, are all too common in the public sector today. According to [Verizon's 2019 Data Breach Investigations Report](#), the public sector faced 23,399 cybersecurity incidents in 2018. Of those attacks, 330 resulted in confirmed data breaches. The government is facing a cybersecurity crisis, with threats ranging from espionage to financial motivation to internal accidents leaving information at risk, including PII and confidential materials.

Traditional cybersecurity measures, such as firewalls and secure passwords, no longer hold up to the evolving threat landscape. Data-sharing and mobility are expanding agencies' perimeters, leaving more gaps for ill-intentioned actors to enter through. Put simply, federal cybersecurity must remain on the offensive by embracing new technologies to overcome this losing battle.

With AI and ML, agencies are able to review data in bulk to actively monitor for the first signs of an attack and respond quickly. AI and ML can give agencies the ability to recognize patterns over time of non-threatening activities to more quickly detect actions that seem out of the ordinary.

This solution provides real-time analysis of the security perimeter, instead of the months or years that it usually takes for government agencies to detect an attack. The [2015 data breach](#) stemming from the Office of Personnel Management (OPM), for example, compromised more than 20 million records and took five months to discover.

After analyzing the structure and common signs of past cyberattacks and attempts based on agencies' data, AI and ML platforms can continuously monitor for new threats. These technologies do not have the same limits as human stamina and are able to audit vast amounts of data for review.

This technology is already being widely used to sort email inboxes. Phishing schemes, for example, are a common way for actors with harmful intents to breach a security framework because it only takes one wrong click from an agency employee to gain access. This style of attack that is so often dependent on human error can now be stopped with the help of large-scale analytics and automation.

Most email providers now redirect phishing attempts to different folders or warn the user that an email contains signs of an attack. This is scalable to fit whatever platform employees are using to access their inbox. With AI and ML technologies, this same idea can be applied to the entire perimeter to take some of the onus off of the staff.

Census Bureau Uses AI to Curtail Data Imputations

Securing good, clean data to make decisions at the federal level is increasingly challenging considering the manual labor and human error that contribute to faulty calculations. Randall Knol, an IT specialist at the Census Bureau's Application Support and Innovation Branch of Demographic Systems Division, shared with GovLoop how his agency is implementing AI to improve data necessary to government decision-making.

GovLoop: How does your team use AI to meet the Census Bureau's mission?

Knol: We look at datasets, and they can be in a large number of different formats, and we profile them to evaluate whether or not they have the data in them that would make it useful to join to our existing datasets that would either extend it, allow us to fill missing data, or check other data for its reliability. With AI and ML, we can change the weight of our evaluation depending on how we assess the accuracy and the value of the data.

And why is having such clean, correct data important to the Census Bureau's mission?

It's the heart of the census. There're the political indications of the way voting in Congress is distributed and the way that government aid is distributed across the country, that's why it's important for everybody to respond to the census. It's how the government makes decisions about the economy. If data's bad, then the government could be wasting time and money on policies that are not productive.

We pride ourselves on having the best data. That makes it a little slower and a little more expensive to accumulate, but our most important attribute is the quality of our data. If AI can help us do that, and not only make it faster and cheaper, but actually improve the quality of the data, then that's really important to us.

How recently was AI implemented into the way that your team does things?

We're just getting started using machine learning and AI. Where it's first being implemented is on the research side. The census is the repository of the administrative data for many government agencies, not just the census data. It's made available under very strict guidelines to researchers. That's where the real usage is being made now. People are using machine learning to speed up their research with a goal of once this research is done and validated, then it will actually move into the actual census surveys.

How does your agency manage large datasets, both from the analytical and storage standpoint?

If you want good data, you have to have good metadata. Metadata is the data about data. Because it does no good for me to have a dataset if I don't know what's in it. We're developing methodologies that, using AI, allow us to generate the metadata from our many datasets. In the past, metadata has been a manual process. It's very slow when it's manual and it's very, very expensive. AI is helping make all these things faster and cheaper. Metadata can change as the data changes. AI makes it faster. We're able to update our metadata and make it more accurate in a quicker way.

Why is AI's ability to speed up the data analytics process important to the work that you do with the census?

Data needs to be good, but it also needs to be timely. If it takes me a year to produce the information you need to make a decision at the end of this month, that's not very valuable for policymaking. I not only have to give you good data, but I have to give it to you in the window where it's viable for actually influencing decision-making.

How is AI helping your agency produce valuable, error-free data?

Most errors come from data entry. The more questions you have people answer manually, the higher your error rate is going to be. The more we can reduce the number of questions we have to ask people, the lower our error rate is and the faster the process is.

What do you see for the future of AI in your department?

Aside from reducing costs and increasing accuracy, I'm hoping that we'll be able to make use of classic big data that comes from the web. It's really not a lot of use to us right now because the data quality is so low, the accuracy is so low and you have to do so much work to clean it up. If we can get AI to do more of this work for us, we can make this much more accurate.

It will never be a completely automated system. But we will be able to focus our trained analyst resources on the really challenging questions. At Census, there's a real big concern about people seeing other people's data. If you've got the AI doing it, then you don't have a person looking at your stuff. It provides a level of privacy and security for people.



CLOUDERA

Insights are hiding everywhere. Manage them anywhere.

Cloudera provides data insight to:

- Identify network vulnerabilities
- Protect sensitive information
- Combat waste, fraud and abuse of government resources.

To find out more,
visit cloudera.com

Holistic, Open-Source Cybersecurity Platforms Are Government's Precursor to AI and ML

An interview with Marcus Waineo, Chief Technology Officer, and Simon Elliston Ball, Product Manager, Cloudera

From the rise of ransomware attacks to large-scale data breaches, cybersecurity (or lack thereof) is dominating government technology conversations. Threats and malicious actors, both internal and external, have the ability to outmaneuver traditional cybersecurity forces, leaving government agencies susceptible to attack.

While time and financial constraints still weigh heavily on government resources, making the switch to a holistic, open-source approach to cybersecurity can prepare an agency to adapt to the evolving threat landscape more quickly and with greater effect.

To learn more about this approach, GovLoop spoke with Chief Technology Officer Marcus Waineo and Product Manager for Cybersecurity Simon Elliston Ball of Cloudera, a cloud-based data management enterprise.

"We want to help government agencies enhance their thinking about cybersecurity as more than a product-based solution," Waineo said. "Cloudera provides a platform that allows you to integrate and optimize the existing tools that you're familiar with into a more agile platform that enables advanced capabilities like machine learning to be applied to the problem."

Elliston Ball explained that the Cloudera Cybersecurity Platform's open-source and holistic approach to cybersecurity consists of pulling together all of the existing tools into a single platform to combine existing data and resources for increased effectiveness.

"There's huge value in just the combination of data that a common platform provides," he said. "Many of the people who are starting out get a surprising amount of value and insight from the relatively simple end of the data engineering side, before they even get on to the AI and machine learning."

AI and ML tools are only as good as the data they're based on. Before an agency can integrate those technologies into its cybersecurity platform, it's important to consolidate, rationalize and cleanse existing data. It's even better when the agency finds a platform to assist in aggregating the existing systems.

Open-source cybersecurity platforms are built on trust and transparency from a broad base of experienced resources from multiple industries and disciplines. This provides a more comprehensive and reliable infrastructure upon which to build a more agile cyber solution. Bad actors share intelligence data within their own circles, but with limited input and data sources.

"The most important part of what we do is to help people make the mind shift away from traditional rule-based cybersecurity and toward a more data science-driven approach." - Elliston Ball

By leveraging the same techniques for good, and with the added advantage of broader and more complete data sets available to the Cybersecurity Operations Center (SOC), Cloudera provides a safeguarded community of trust for government agencies to do the same. The Cloudera platform delivers the performance, scalability and reliability needed for agencies to efficiently organize and manage big datasets and advanced analytics more effectively.

With this platform, data scientists can safely and swiftly deploy AI and ML cybersecurity models with greater effect by bringing their tools directly to the data. Holistic, open-source data management removes analytic silos to drive more value from the data an agency already has.

Takeaway: Bringing data together in open-source platforms provides a comprehensive understanding of the government agency's network to help detect and thwart cybersecurity attacks.

DHS Embraces AI for the Future of Cybersecurity

At the Department of Homeland Security, the mission is simple: Keep the nation secure from all of the different threats it faces. But a lot has changed since it was created in 2002, including the spread of security technology at its disposal. Martin Stanley, Senior Advisor for AI in the Office of the Chief Technology Officer at DHS, told GovLoop about the agency's current AI-led cybersecurity efforts and shared his team's best practices for applying AI to its existing framework.

GovLoop: What were some of your agency's initial use cases of AI and ML to combat cyberattacks?

Stanley: AI and machine learning have had a long history with cybersecurity. The application of those technologies is not new. Those technologies have been pretty widely deployed, and we're already using them in our programs that we have here, such as CDM [Continuous Diagnostics and Mitigation] and DPS [Defense Personal Property System]. We do anticipate an increased use of these approaches through other applications like incident triage and Security Orchestration Automation and Response [SOAR].

Tell me more about those applications.

Incident triage is really challenging because you have tons of data about what's going on in your environment coming in all the time, and it's increasing as there's more sensors and there's more attacks. Having an automated capability to review this information and to pull out the ones that should be responded to, or the ones that should be referred to a human agent, is a big application that we see both within our mission states here at DHS, and also with the communities and stakeholders that we serve.

Security Orchestration Automation and Response is a capability of taking a predefined set of actions based on the analysis that a machine is doing in the environment. It sees a certain set of information coming in through sensors, and it makes certain responses in an automated way.

Could you give an example of a specific project or program that's teaming together humans and AI capabilities to improve cybersecurity?

The Continuous Diagnostics and Mitigation Program is a civilian, agencywide program where we deploy sensors and a dashboard reporting system across the federal government to identify cybersecurity issues to report those back up through the federal dashboard, and then also to send threat information and recommended responses back down to the federal civilian agencies. Within that, there's a lot of opportunity for automation at the agency level where they're doing their cyberdefense.

The EINSTEIN system is a perimeter protection system for federal civilian agencies. With all this data that we have about the perimeters of federal government, being able to do an automated analysis of that information so that we can prioritize the response actions and the incidents for the human operators is an application that we're looking at as well.

What types of AI solutions are you exploring, more generally?

We're looking at narrow AI solutions, which are good at uncomplicated, known tasks that have tons of data examples. We focus on what the best practices for implementing AI or machine learning systems within our environment are and then try to apply those as we look at solutions to see if they make sense.

What are some of those best practices?

Make sure that you look at what data you need in order to train the algorithm in how it needs to be prepared. You want to look really hard at the data that you have to see if you've got the right data and you have enough of it that you can sufficiently utilize an algorithm.

Another best practice that we've identified is to have a human performance metric that you're using to measure how well the AI is performing the task. Use that as your benchmark to determine if your AI doing better or worse than your human.

The three dimensions that you're really looking at are benefit, regret and complexity. You don't want to deploy a solution that's going to be of low benefit. You want to have low regret in the event that the machine doesn't do what you want it to do, or the machine makes a mistake. And then lastly, there's a complexity dimension.

These machines generally tend to not be very good at high-complexity tasks, but they tend to outperform humans at low-complexity tasks.

What do you see for the future of the relationship between AI and cybersecurity either within your agency or at large or both?

We're concerned about AI from three perspectives. Obviously, we spend a lot of time talking about how AI can be used to perform our cybermissions within our program space. But one of the areas that we're also concerned about is how the networks that we protect are going to deploy AI, and how that's going to change the attack surface that we're helping to protect. AI creates additional vulnerabilities in an environment based on the inherent nature of these systems. Lastly, we're concerned about how adversaries are going to use AI to conduct attacks.



Conclusion & Next Steps

Using AI and ML to combat cyberattacks and fraud will look different for every agency. That said, there are a few tangible steps agencies can take to set the groundwork for the new technology. These suggestions will make sure your agency has the right framework in place to work AI and ML technology into its daily cybersecurity functions.

1. Identify a specific problem that can be solved or goal that can be met using AI or ML.

Aligning your agency's mission with the AI cybersecurity or anti-fraud strategy can help get the entire agency onboard with implementing the new technology. A definitive goal to work toward or problem to address helps define the AI strategy and keep the team on track.

2. Modernize legacy IT systems.

Before implementing AI in any capacity, including to prevent cyberattacks or fraud, your agency must ensure that its current systems can handle the new technology. Oftentimes, legacy IT infrastructure cannot support the large datasets and heavy traffic necessary for effective AI or ML strategies.

3. Improve data available.

Data can be messy and sporadic. For AI and ML technology to properly analyze the necessary information, the data must be clean and organized. Break down the silos that separate data for easier access with a holistic data management platform or by improving internal data organization.

4. Break down management silos.

The silos that separate an agency's development, security and operations teams must also be reworked for better collaboration. Consider a DevSecOps approach or a unified management platform to foster collaboration and innovation on this front.

5. Provide training on AI fundamentals and agency-specific use.

Your team must understand how AI works, where it will fit in the cybersecurity framework and your plan for implementation. This ensures an understanding of the big picture and the individual roles in meeting goals along the way.

6. Establish a way to track success.

Based on your goals, consider what success will look like as your team progresses. While the end goal may be clear, embrace the small wins to encourage progress.

7. Build a cybersecurity model around the available data and infrastructure.

Once the staff and IT infrastructure are ready to implement AI and ML into the cybersecurity framework, your agency is ready to build a model that attempts to meet the specific goal that the team set out for. The model should rely on the data and infrastructure that your agency currently has in order to optimize existing resources.

8. Test the model and analyze results.

Before rolling the model out, it's important to test for possible flaws. Implementing new cybersecurity tools has huge stakes and you'll want to ensure that your agency is not left susceptible throughout the transition. Test rigorously and enhance data to keep improving.

CLOUDERA



Thank you to Cloudera for their support of this valuable resource for public sector professionals.

About GovLoop

GovLoop's mission is to inspire public sector professionals by serving as the knowledge network for government. GovLoop connects more than 300,000 members, fostering cross-government collaboration, solving common problems and advancing government careers. GovLoop is headquartered in Washington, D.C., with a team of dedicated professionals who share a commitment to the public sector.

For more information about this report, please reach out to info@govloop.com.

govloop.com | [@govloop](https://twitter.com/govloop)