How Data Drives Innovation in Government

NetApp®   TechData   govloop

Executive Summary

By now, you've heard all the talk about making government "smarter" — that is, making agency operations and services more efficient, effective and responsive to changing requirements. But how do governments achieve that? One way is leveraging the power of emerging technologies and using artificial intelligence (AI) to its full potential, but that's easier said than done.

The potential for transformation is stunning, but let's remember the fundamentals first. What goes into a smart government is much more than an acquisition. Smart governments come from organizations that do the little things right and put the building blocks in place gradually.

The first block is data. When data is transformed into useful, meaningful information, it becomes one of the most important public assets. To create the transformational solutions that drive a truly smart and innovative government, agencies must ensure that data can be easily accessed, efficiently stored and securely protected.

Nowhere is that truer than with AI. Agencies are just beginning to explore the many ways in which they can use AI to improve their operations and develop a new generation of services. But they are also discovering that they cannot get far with AI without a strong foundation of accessing and extracting value from their data.

In this e-book, "How Data Drives Innovation in Government," you'll learn how smart governments use their data to power world-changing innovations, with a look at both the opportunities for leveraging data and the issues that you will want to address to make it a reality.

Data-Driven Government at a Glance

Statistics

- AI can drive automation that frees nearly **97 million** government working hours each year, saving agencies **$3.3 billion annually**.

- The federal IT spending priorities will focus on investing about **$4.9 billion** in unclassified AI and machine learning-related research and development in fiscal 2020.

- The Army announced that it was investing **$72 million** in a five-year effort to research and discover capabilities that would significantly enhance mission effectiveness by augmenting soldiers, optimizing operations, increasing readiness and reducing casualties.

- *"The federal government can help the U.S. maintain its leadership in AI by working closely with our experts in industry and academia, investing in research, and engaging with the international standards community,"* said Walter Copan at NIST.

- By 2025, there will be more than **26 smart cities** worldwide.

- The global smart city market is expected to reach a value of **$1.57 trillion** by 2020.

Recent federal initiatives

- Federal Chief Information Officer Suzette Kent officially unveiled the **Federal Data Strategy** in June 2019. It includes principles and practices for federal agencies' data governance along the lines of mission, service and stewardship. "What you will see is, like any lofty goal, we are looking both strategically and tactically, and we have to start with the basics and invest and build a rock-solid foundation," Kent said. "And the framework that we're sharing today supports raising the bar for consistency of skills, interoperability and the best practices for how the agencies manage and use data." The strategy itself is broken into three categories of practice around culture, governance and promotion.

- In February 2019, President Trump issued an **Executive Order on Maintaining American Leadership in Artificial Intelligence**. It lists five principles that will guide the new program:

  1. Driving AI breakthroughs in the United States across the federal government, academia and industry to promote economic competitiveness, national security and scientific discovery.
  2. Developing technical standards and reducing barriers to the safe testing and deployment of AI for today's industries and future ones.
  3. Training current and future U.S. workers so that they have the skills to develop and apply AI for both the current economy and tomorrow's version.
  4. Fostering public trust and confidence in AI so that the technology is used to protect civil liberties, privacy and American values.
  5. Creating a global environment that supports American AI innovation, research and open markets while maintaining the nation's edge on the technology. The United States must also protect its critical AI technologies from adversarial nations and strategic competitors.

- NIST has released a **plan for prioritizing federal agency engagement** in the development of standards for AI per the executive order. The plan recommends the federal government "commit to deeper, consistent, long-term engagement in AI standards development activities to help the United States to speed the pace of reliable, robust, and trustworthy AI technology development."

Defining emerging tech

New data analytics tactics are changing the way that government does its job, including AI, machine learning and robotic process automation, or RPA. What are each of these?

AI: AI enables computers to mimic human behavior. It's achieved when a system mimics human thinking, reasoning and decision-making by following logic and rules explicitly programmed by humans.

Machine learning: A system uses data and experiences to make itself smarter over time. It's not separate from AI; it elevates AI from a system that follows instructions to one that actually knows how to learn without needing additional explicit programming by humans.

RPA: RPA enables you to create software robots, or bots, to automate business processes. Using technology that mimics human behavior to interact with existing human/system interfaces, RPA essentially allows for the creation of a virtual digital workforce capable of performing simple to complex repetitive data entry or data movement tasks and business processes.

The Potential of Data

Data can power innovation in government. Here are two examples that show how agencies can use the combined power of AI and datasets to create, innovate and better serve constituents.

Case Study: AI Powering Acquisition

Purchasing new solutions and contracts in government can take years. Although it seems like spending money in the public sector should be as simple as clicking "add to cart" like businesses and people in general do, governments undergo intense security checks, systems integrations and lengthy schedules for implementation. A solution can turn from innovative to archaic throughout the course of a procurement process.

But emerging technologies such as AI and blockchain provide the means to accelerate the delivery of new capabilities, especially in the area of automation.

The Department of Health and Human Services is a case in point. HHS has automated key acquisition processes through a combination of AI and blockchain, a distributed ledger technology that enables digital information to be shared in a way that it cannot be changed without authorization.

AI and blockchain have been incorporated into Accelerate, the department's acquisition portal. The platform's database of contracting data from the agency's five acquisition systems will drive automated pricing breaks for government the same way consumers can use competitive data at major retailers.

Although AI and blockchain drive the automation, the program hinges on good data management practices.

"We believe this infrastructure is the way government will modernize all of its administrative functions and then extend itself with those savings into the mission space," HHS CIO Jose Arrieta told GovLoop.

Use Case: Enabling Improved Citizen Interactions with AI

Today, with the rapid speed of technological innovation, in addition to rising citizen expectations, agencies are looking for new ways to increase citizen engagement and satisfaction. Keeping pace with citizen experience expectations has become even more complex because of the fast growth of channels and interactions. Increasingly, AI is seen as key to a solution.

According to a General Services Administration whitepaper, the accelerated adoption of emerging technologies such as AI, RPA and the Internet of Things are critical to the future of the government center. These technologies "will take on routine aspects of contact center roles (e.g., answering frequently asked questions) leaving the non-routine (e.g., addressing more complex customer inquiries) to humans."

By 2022, 70% of customer interactions will involve technology such as AI and machine learning, GSA has said. As this future unfolds, agencies will be able to look to the private sector for best practices in using AI to improve customer experience (CX).

Anyone who has been to a government field office to address an issue realizes that government workers and citizens alike would greatly benefit from advances such as an AI-driven self-service kiosk, similar to those used in commercial services, to pair citizens with the appropriate representative or service – and in the process, free up time for employees to take on more complex tasks.

The same is true online. Using AI-based systems to expedite case processing and CX requests to ease bottlenecks and by providing more self-service options, such as chatbots and online portals, the government can give constituents quality assistance faster and without making them wait on hold or at field offices.

Again, data is key. The federal government touches the lives of almost every constituent and thus has an immense amount of data to use to create personalized and efficient services. The challenge is putting in place the right systems and processes to ensure that the AI systems can use that data.

A Matter of Trust

The value of data depends not just on its volume or content but on its trustworthiness. Trust is especially important when it comes to using AI, because if there are any questions about the underlying data AI, the resulting insights or operations will be called into question as well.

How can your agency make sure trust is built into the data and the AI tools you are using? In a recent report, NIST identified eight tools that agencies can use as a checklist on their journey to leveraging data for AI and innovation:

1. **Standardize dataset formats, including metadata for training, validation and testing of AI systems.** Data standards are vital for measuring and sharing information relating to the quality, utility and access of datasets.

2. **Promote consistent formulation of, reasoning with and sharing of knowledge** with tools for capturing and representing knowledge and reasoning in AI systems, thereby promoting interoperability of AI systems and minimizing their misunderstandings and inferential errors.

3. **Study fully documented use cases** that provide a range of data and information about specific applications of AI technologies and any standards or best practice guides used in making decisions about the deployment of these applications.

4. **Use benchmarks, evaluations and challenge problems** to drive innovation by promoting advancements aimed at addressing strategically selected scenarios.

5. **Use testing methodologies** to validate and evaluate AI technologies' performance, especially to prescribe protocols and procedures.

6. **Use metrics to quantifiably measure and characterize AI technologies**, including but not limited to aspects of hardware and its performance, trustworthiness, complexity, domain-specific and context-dependent risk, and uncertainty.

7. **Implement AI testbeds** "so that researchers can use actual operational data to model and run experiments on real-world system[s]…and scenarios in good test environments."

8. **Use tools for accountability and auditing** to enable examination of an AI system's output (e.g., decision-making or prediction). These tools can improve traceability by providing a record of events and information regarding technologies' implementation and testing. In so doing, they can enhance assessment and documentation of gaps between predicted and achieved AI systems' outcomes.

Optimizing Your Data with the Power of AI

Edward Dowgiallo, Architect at the Transportation Department's Office of the Chief Information Officer and Federal Transit Administration's (FTA) Office of Information Technology, is helping deploy AI for pattern recognition at the agency. At a recent GovLoop Virtual Summit, we spoke to him about how AI is helping with datasets on everything from public transit to road safety.

This interview has been lightly edited for clarity and length.

GOVLOOP: Can you explain how data optimization is laying the groundwork for AI at your agency?

EDWARD DOWGIALLO: AI is not yet at a point where it can replace people and the jobs they do. It's more specific to tasks that people can do. It can help automate those tasks, but when I look at AI, what it helps do is identify patterns in my data that I can use in a repeatable manner. Given that AI's about pattern recognition in your data, if you have optimized data, it would naturally make sense that you would be able to do some of those activities more efficiently. And so, in my agency, the thing we're very invested in is really building a data capability so that we can eventually build machine learning algorithms.

How is AI helping with the day-to-day work being done at FTA?

We started with AI around something we call the National Transit Database, which collects transit statistics for the whole country on public ridership. We were really concerned about the quality of the data, so we started paying people to do data analysis to make sure that information was accurate, and then we started building a rules-based system, to make sure that that data was clean. As you know now, with AI or machine learning, what you can do instead is you can make the data more accurate.

Another area AI is helping us with our data is safety. We want to start benchmarking our data and understand situations where, for example, if we drop maintenance for a transit agency by $1 million, do accidents increase.

Finally, we want to use AI to understand our data well enough eventually to be able to do predictions on transportation itself, and be able to say, as an example, "If I put a bus route on 14th Street in D.C., that will provide X percentage increase of accessibility to jobs for the people that live there." So those are the types of use cases we're doing or looking at potentially implementing with AI.

There's a lot that goes into determining who is involved with agencies' AI efforts. How do you decide who is part of those conversations?

Having the business lines involved is very important. Since I work in the IT shop, with the data, you need somebody that's a topic expert as well. At my agency I'm happy to see that we're starting to have data scientists incorporated into our business lines. And we need more of those to exist throughout because there needs to be an expert or an owner of that data. For example, when I'm dealing with the National Transit Database, it helps that I have somebody from that business line that can look at the patterns that we're recognizing in the data and they can say whether or not they think there truly is causality there, or if it's really false positives or inconclusive things.

AI is something that augments the human, but the human still needs to make the decision. And the human needs to give this an eye test. So, we need some sort of expert that can take the contexts or the indicators that the machine learning algorithms are giving us, and interpret it – determine whether they think it's accurate, or point to other things that we should be looking at to explain why we are seeing correlations or causality.

Understanding Data as Source Code

David Arnette is the Senior Technical Marketing Engineer at NetApp, which provides a full range of hybrid cloud data services that are designed to accelerate digital transformation. Arnette is focused on solutions for AI and machine learning workloads. At a recent GovLoop Virtual Summit, we spoke to him about how agencies can provide a solid data foundation for their AI programs.

This interview has been lightly edited for clarity and length.

GOVLOOP: How does data optimization help lay the groundwork for AI?

DAVID ARNETTE: One of the things we like to say is that data is the source code for AI. Typically, in software development, you have people writing software that makes things happen. But in the AI space, it's almost reversed: The software is doing the writing, the software is learning to identify things, or classify transactions, based on its analysis of the data. So, the quality and the accessibility of the data are absolutely critical to success with these types of initiatives.

How might agencies figure out what problems they can solve with AI and machine learning?

The first things I usually see customers look for are the low-hanging fruit – things that can deliver big impact without a lot of effort. In a lot of ways, that's actually tied to the data that's available at any given time. But when efforts are starting up and people are learning how to execute this type of program, they realize that they don't have enough data or can't get to it in a timely manner. That data exists, but it's in a variety of different places and a variety of different formats. The key is to make sure that the right pieces of data are being collected in a centralized manner, so that they can be acted on holistically.

How do you create a data governance model that will aid AI? And to what extent is that model from agencies traditionally used?

For AI initiatives, we see a lot of need for version control of the data, in addition to the code itself. This is a very collaborative environment, in which datasets often get combined. Once the data scientist has created a good dataset, it's very useful to be able to share that dataset and provide a well-understood "access package" spelling out permissions for the team that might use it. And then that follows through all the way into archiving and cold storage of the data. At some point, we may need to reference the specific datasets that were used to train systems, so that data needs to be protected for longer periods of time and with more control and understanding of how that data relates to the other datasets that may be in use.

How do you see agencies moving forward with data optimization and AI in the next year? What's on the horizon?

I think everybody is starting to realize the importance of data. It's gone through phases where, first, there was not enough of it, then there was too much, and now it's becoming an accepted fact that the only way forward is to maintain and manage data. The other issue facing a lot of organizations is that data is spread across many physical and logical locations. Managing the data across all of that will be key to success in almost every organization.

Improving Trust with AI at U.S. Customs and Border Protection

*When your agency produces 1.3 petabytes of data per year, how do you make sense of it all? At a recent GovLoop Virtual Summit, we spoke to **Neil Chaudhry**, Director of Regulatory Audit Systems and Innovation in the Office of Trade for U.S. Customs and Border Protection, about how AI is helping sort all that data.*

This interview has been lightly edited for clarity and length.

GOVLOOP: What is your perspective on how data optimization is laying the groundwork for AI?

NEIL CHAUDHRY: My office processes $2.6 trillion in imports, generates about $42 billion in revenue for the U.S. government, has 35 million declarations and generates about 1.3 petabytes of data a year. On a daily basis in our work, we have a tremendous amount of data. I'm part of the audit portion of the organization, where we make sure that the importers, the ecosystem as a whole, have trust in the system, that everybody is playing by the rules, everybody is doing the right thing, and that bad actors don't have an advantage over people that are following the rules. So having the data in a clean format, a structured format that's easily accessible and accessible quickly, is critical for the type of work that we do.

AI helps us focus on improving the effectiveness of the audits that we do and helps us deal with the data as we continue to scale and build more trust.

What's an example of the kind of work that you are doing with AI?

AI helps us understand the predictive impact of things like presidential proclamations or new regulations that may come out. What is the impact of them? Would it limit the imports coming in from a particular country? If we increase the tariffs on a particular commodity, does the way that the importer is classified, does that commodity change? And using AI helps us build better models. AI lets us run huge datasets and start deducing things we couldn't otherwise with machine

A snapshot of CBP:

$2.6 trillion
in imports

$42 billion
in revenue

35 million
declarations

1.3 petabytes
of data per year

learning models. Our machine learning and AI journey is about augmenting all the good work that our current subject-matter experts are doing. And clean data is so important both for AI and our experts. When our models are not working the way they should, it's often because we're not collecting the correct type of data, or our data is too corrupt to draw any meaningful conclusions. Those are things that we continue to work through.

Do you have advice for how agencies can get started on their AI journey?

In government, we often get inundated with nice, shiny, fancy tools, or go to a presentation and there's this razzle and dazzle. But when we come back to our agency, we find that it's not there – we don't have the infrastructure, the workforce, the capability or the capacity to use those tools. So I always say when you start your AI journey, start with small sets of data, small problems and simple programming tools.

There's absolutely nothing wrong with starting with a small dataset and trying it out. It should be a project that's not on

somebody's performance plan, so that even if it doesn't work out, there's institutional learning that occurred.

Finally, joining communities of interest is a great place to start. GSA has a number of communities of interest. Email somebody to ask for a quick bit of advice on something. I've leaned on people, I've asked for advice and I'm always happy to give advice. Build those networks to start the AI journey because it starts with the analyst. Of course, you have cover, and you have encouragement from the leadership, but the leadership isn't there writing the code, building the models. It's actually the rank and file doing that.

Conclusion: Is AI in Your Future?

There's no doubt about it: In the next several years, AI's role in government will only grow. Ultimately, AI will enable agencies to increase both the efficiency and effectiveness of their operations and deliver services that are more flexible, scalable and responsive to changing requirements. In short, AI will make agencies smarter.

The task now is to provide a solid foundation for that future by focusing on the data that fuels AI. Agencies need to put in place data management systems and strategies that ensure that data is well-organized, well-scrubbed and analytics-ready.

The good news is that the disciplines and tools around data management have matured significantly in recent years, thanks to the accelerated interest in big data. That trend recently culminated in the release of a Federal Data Strategy.

In any case, the work that agencies do now to get their data in order will pay off many times over in the years ahead.

NetApp®

TechData

Thank you to NetApp and TechData for their support of this valuable resource for public sector professionals.

govloop

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