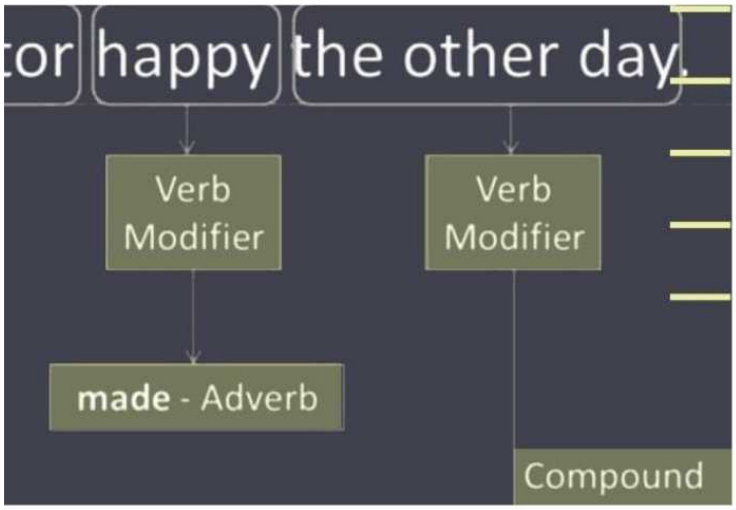
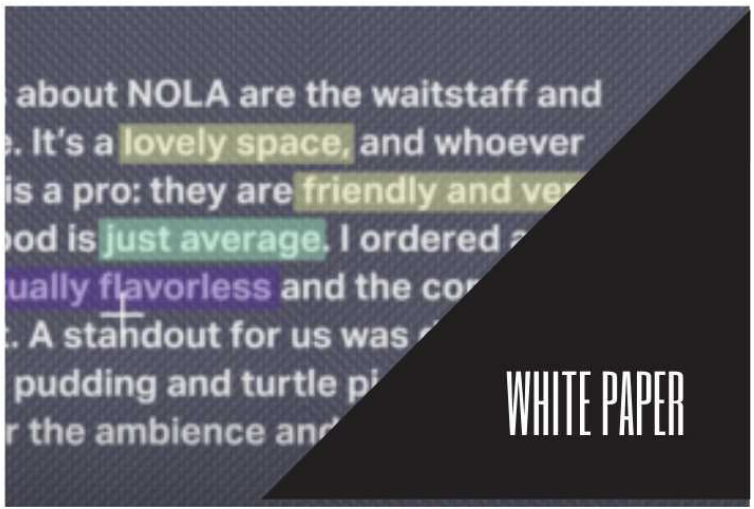
iMerit

GOVERNING
WITH (ARTIFICIAL) INTELLIGENCE





BIG DATA AND THE RED QUEEN EFFECT

"In the 21st Century, data will eclipse both land and machinery as the most important asset." So says futurist and social scientist Yuval Noah Harari, in his book, "21 Lessons for the 21st Century."

Indeed, it is hard to overstate the pervasive influence data can wield in our lives. When harnessed for the good of the populace, government data contains the potency to vastly improve the quality of life for citizens and the quality of work for government employees. The plenitude of data is staggering. By 2025, research firm IDC estimates all the data that we create, capture and replicate will total 175 zettabytes -- or 175 trillion gigabytes. One government stakeholder who agrees with Harari on the epoch-making value of data compares it to "the iron and ore of the industrial age." The technology that makes data a profound game-changer for the presumed betterment of society is Artificial Intelligence (AI).

We've all heard of AI, but what is it exactly? Let's go to the source. That would be Stanford University professor John McCarthy, who first used the term in 1955, describing Artificial Intelligence as the "science and engineering of making intelligent machines, especially intelligent computer programs." It is the limitless talents of those intelligent computer programs that invest data with virtually superhuman capabilities -- to summarize, to analyze, to prescribe solutions, to predict and to prevent outcomes. Through AI, we can teach machines -- to emulate and understand human language and speech; human sight; and human intelligence -- so they in turn can teach us.

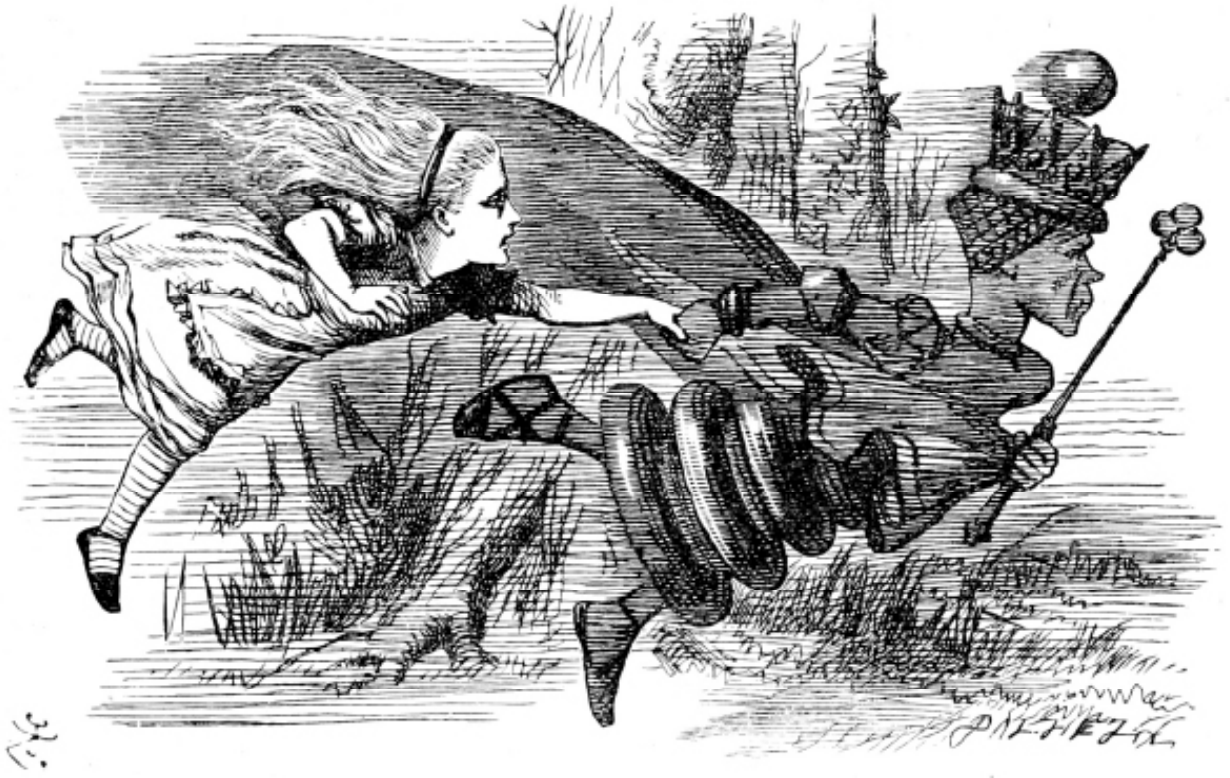
In more technical terms, those are known as the AI modes of Natural Language Processing (NLP) and Computer Vision (CV) -- the latter also known as Video Analytics (VA). They are in turn enabled by Machine Learning (ML). Using linear regression, Machine Learning is a subset of Artificial Intelligence that empowers unstructured or structured datasets, which are either built or imported, to recognize patterns

and to steer the ML client to specific decisions or insights. Machine Learning is the teacher that trains AI which patterns to look for, so AI can respond when it recognizes them. ML, mostly used currently for prediction and classification, has opened a whole new world for the normally glacial-moving machinery of the public sector.

In government circles, they talk of the Red Queen Effect. It's an experience relatable to virtually anyone who's worked in a middle management or bureaucratic capacity, whether in the public or private sector. The Red Queen Effect posits that the daily grind of a job function -- especially a rote job function -- creates a kind of inertia that inhibits innovation. A numbingly repetitive vocational routine all but renders innovation literally unthinkable.

The complex infrastructure of government faces a kind of dilemma -- a double-edged sword, if you will. It has at its disposal vast amounts of data from which it can derive an equally vast knowledge base of information, insights, and solutions to better serve humankind. How are government workers, though, expected to transform all that data into actionable insights and viable solutions? It is painstaking work that requires highly specialized skillsets and an enormous amount of time. It certainly is not practical with legacy office workflow such as spreadsheets.

Civil service workers will readily agree with what the Queen tells Alice in Lewis Carroll's timeless parable: it "takes all the running you can do just to keep in the same place." If government wants to fully unleash the magical powers of the genie inside data, it must "go somewhere else," to further quote the Queen, who advises Alice to "run at least twice as fast." Enter the blinding speed and limitless potential of Artificial Intelligence (AI), which continues to gain traction in all corners of government and in all corners of the globe.



More than four out of five organizations in the public sector intend to use AI, according to a Deloitte study. As use cases illustrated in this paper attest, government departments are making headway leveraging some form of AI, typically for the sake of efficiency as well as improved efficacy and accuracy in outcomes that serve constituents. How agencies use AI and its modes tends to depend on whether they are more oriented towards research and development or towards policy-making. For example, the research-focused NIH (National Institutes of Health) has an AI budget ($30 billion) more than double that of the policy-centric DoL (Department of Labor).

Department budget policies provide a window into their relative affinity for AI adoption. The Centers of Excellence (CoE) unit within GSA (General Services Administration) -- which includes an AI subdivision -- recommends agencies allocate the 20% of their budget not used for day-to-day operations to modernization. Of that, 15% should be earmarked for upgrading systems and 5% for innovation. What ends up happening in too many cases, however, is that the 20% put aside for "futurizing" the agency is effectively folded into the maintenance of the same, old status quo. Another AI challenge for government is that "government units don't know what data they have," says Infor VP Bob Benstead. More optimistic is the survey that says nearly 70% of government leaders agree that there is value in upgrading their data analytics platform. A logical entry point for government functionaries to apply AI are the basic but very versatile Machine Learning capabilities of NLP and CV.

The ingenious ability of NLP to find patterns in otherwise abstract data is being deployed at agencies to improve the speed and quality of decision-making, upgrade services, and protect populations at risk, whether due to environmental hazards, health crises, cyber attacks, economic vulnerability, or terroristic threats. In that way, government units can study patterns on matters of public interest, and more rapidly organize mountains of data for optimal efficiency.

FEMA (Federal Emergency Management Agency) uses enterprise analytics to aid people before, during and after such disasters as Hurricane Sandy in 2013. The data helps in the research and delivery of preparedness grants, training and exercises, and pre-disaster mitigation. Other data-driven activities are coordinating the federal response, managing resources, and strategizing long-time recovery.

Australia's National Criminal Intelligence System uses ML to collect and monitor data on criminal activity so it can prevent anti-social activities instead of always having to respond after the fact.

One reason NLP is a popular use of AI to date in both public and private sectors is because NLP uses as its raw material the massive amounts unstructured data generated by all of us on a daily basis -- through social media, email, phone conversations, web documentation. Yet, 80% of all data is unstructured. That's where the human element of NLP enters the equation, to annotate, label, and enrich. NLP processes the data into valuable insights about human behavior, public sentiment, policymaking, criminal intent, security leaks, and other information that greatly enhances government services and empowers government constituents.

Computer Vision (CV) is another AI technology finding its way into a growing number of government agencies. By monitoring strategically placed cameras in public spaces, CV can be used to anticipate security threats such as flash mobs forming spontaneously with disruptive intent.

It can see and report from still and moving images potentially dangerous objects such as guns, or potentially dangerous public safety threats such as compromised infrastructure in need of imminent reinforcement or replacement. Video Analytics offers increasingly relevant utility too as a high-tech means of compiling and analyzing the footage from body cameras that are fast being adopted as mandatory gear worn by uniformed law enforcement officers.

Career civil servants facing the sunset of their careers are neither conversant in new technology nor interested in adapting to it.

As much progress as has been made in developing use cases for the NLP and Computer Vision modes of AI, there remain formidable barriers and challenges to surmount. While there is no shortage of data, having the financial wherewithal and requisite human resources to extract value from that data remain beyond the grasp of many agencies. New budgets must be created and approved to acquire and process the data. That's not to say all the data is ready for processing. That creates a snag within government agencies, where there is a notable shortage of data scientists with the skillset or training to structure data, a necessary step to make it compatible with the algorithms that are the heart of Machine Learning and Natural Language Processing.

A long-term strategy, that needs to be put in motion "yesterday," as they say, requires the expense of recruiting and training a next-generation workforce. The current average age of government workers is older than the comparable private sector average age. In the public sector, 27% of the workers are millennials, compared with 38% of workers in the private sector who are in that cohort. Career civil servants facing the sunset of their careers are neither conversant in new technology nor interested in adapting to it.

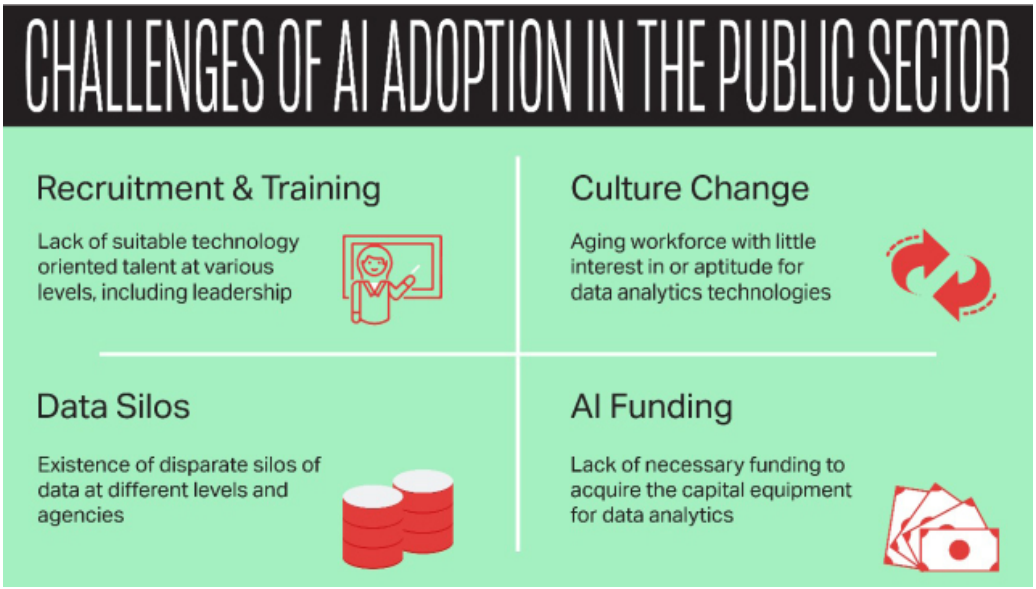
At the local level, the situation can be starker. In Portland, for example, almost one-third of the government workforce is retirement age, reports its Public CIO. There also has been a lack of suitable tech talent at the leadership level, with those policymakers acting effectively as a roadblock to their agencies' adopting valuable capabilities such as NLP and Computer Vision. Deloitte analysts contend that the government must cast a wider net to recruit talent from all corners of the technology sector if it is to get up to speed with AI and exploit to the fullest extent its varied attributes for the benefit of the citizenry and the common welfare.

In the meantime, help is on the way in the person of an increasingly visible role at federal and local levels: the CDO, or Chief Data Officer. It's a critical position charged with a purview of, to put it deceptively simply, "finding solutions to public problems," in the words of a Deloitte-Beeck Center report.

The existence of disparate silos of data at different levels and agencies is another major issue to be addressed in promoting the use of NLP and CV within government groups. The lack of a cohesive, standardized data platform is holding back AI adoption in government, says the Center for Digital Government in a report for the Department of Health and Human Services (HHS). "Any kind of analytics and reporting based on silos will give you a very narrow view of things," according to Anthony Farinha of cloud database company Couchbase.

The Center for Digital Government's senior fellow Otto Doll is more blunt in his assessment of the need to formulate a single system of data processing: "If the public understood what kind of decisions are being made by a department or agency relying solely on the information they collect themselves, they would be shocked." The Center points out that by having a unified platform with a standardized toolkit of analytic models, predictive algorithms will be easier to build, lowering cost and time cycles for managing a public health crisis like the Covid-19 pandemic.

Another significant advantage of a unified platform is that it has the potential to simplify and democratize access to the analytics dashboard. That means even an analyst without high-end data science skills in using NLP still would be able to mine the data for valuable information without having to be data-literate in esoteric jargon or parameters. That kind of open access platform is analogous to someone who doesn't know computer coding having the ability to manage a website through the user-friendly dashboard of a content management system like WordPress. On a macro level, at least, the will appears to be there to do what it takes to turn AI to the government's advantage in all ways possible. A Deloitte survey says more than 80% of public sector entities indicate they plan to use AI and almost 9 out of 10 consider cognitive technologies "of extreme importance" to their business.



Source: iMerit

NLP SPEAKS MACHINE LEARNING'S LANGUAGE

Outside of nuclear energy, there arguably never has been a more transformational technology available to government than the Machine Learning (ML) capability inherent in Artificial Intelligence. The more data that is generated, the more difficult it becomes for government workers to manage it all in a way that best serves their respective populations.

Using the algorithms programmed into ML, Natural Language Processing (NLP) takes text or voice data, analyzes it, detects patterns and anomalies, and delivers meaning and context based on the parameters it has been given by data scientists In this way, the knowledge workers inside government agencies can more quickly and effectively categorize disciplines they oversee, draw conclusions from public sentiment, and plan for the future by predicting likely market patterns.

Data can be thought of as the grey matter of Artificial Intelligence. With the government collecting and generating data measured in zettaabytes, and with hundreds of millions of citizens to enable, serve and protect, it's no stretch -- if a bit corny -- to call government's affinity and utility for Natural Language Processing and Machine Learning a match made in heaven. This is being borne out by the growing range of agencies and other government functions adopting NLP to fulfill missions.

In healthcare, the U.S. National Library of Medicine uses NLP to protect patient privacy by "de-identifying" their confidential medical records. In pharmacology, the National Center for Toxicological Research accessed the archives of the FDA (Food and Drug Administration) for the purpose of predicting dangerous reactions to certain combinations of drugs. In ecology, NLP made it faster and more feasible for the Department of Energy to determine, in order of cleanliness, the top energy innovation ecosystems in the country.

Beyond their ability to read human language and generate and understand text in a literal sense, the value of NLP to the complexity of federal services is greatly enhanced by additional talents. It can perform sentiment analysis to interpert intentions in a human voice or statement, and identify and classify categories of information (entity recognition). That's of tremendous import for understanding the critical issues on the electorate's mind, and for assessing public opinion on those issues.

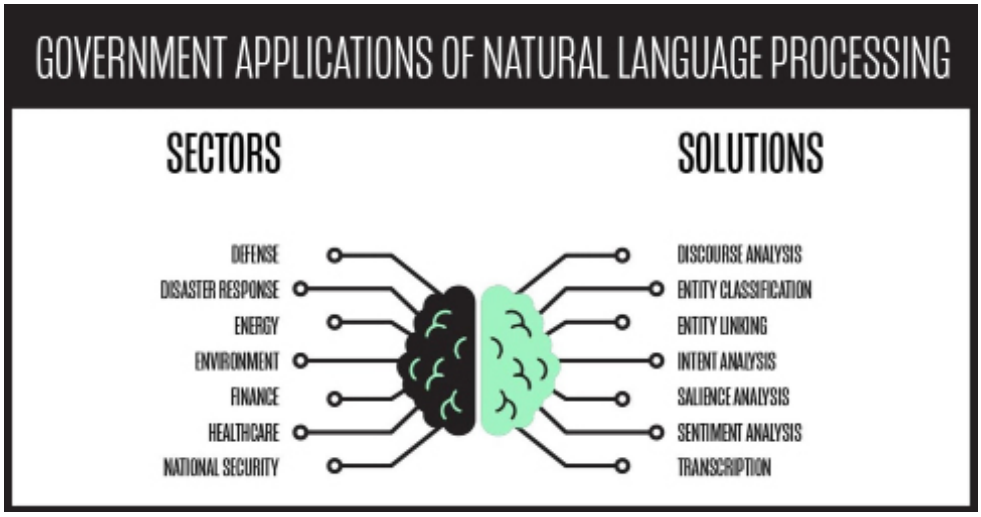
For the algorithm to be trained, though, it can't be fed raw, unstructured data -- data that can come from social media, email, websites, call centers, field reports and many other sources. SInce 80% of data is unstructured, the datasets that are uploaded for NLP must first be labeled, annotated and enriched to respond to algorithms that are programmed to teach it. In that way, classifications and patterns can be detected that are used to train the algorithm for a specific purpose, such as detecting fraud. That process requires data scientists who are expert in analyzing and annotating the data to ensure the enriched data meets the rigorous standards required to produce accurate results.

Machine Learning's two primary functions are to analyze or identify existing patterns and to predict future events or activities. FEMA (Federal Emergency Management Agency) has used ML science to locate shards of infrastructure buried in lava flow. A midwestern municipality used ML to forecast when potholes would appear. The Department of Defense is the leading proponent in the public sector of AI in general, and NLP in particular. In fact, it was the agency's DARPA branch that hatched the concept for Siri, the seminal digital voice assistant.

The DoD's budget for NLP is roughly $100 million. In one hypothetical scenario constructed by Deloitte [article by Eggers etc..Jan. 16, 2019], a Department of Defense team is able to evaluate potential threats and help direct mission strategy by gathering weather data from the National Weather Service, traffic data from the Department of Transportation, public online comments, details on military maneuvers, and social media posts.

Where weather and military activity are examples of structured data, in contrast are the anecdotal content of social media, online conversations, and field reports by the military -- all of which are unstructured. To detect trends from that material through manually studying it is overwhelming in terms of time, energy, and cost. The solution that resulted from this challenge is the work of DoD's DARPA (Defense Advanced Research Projects Agency). It's a program called DEFT (Deep Exploration and Filtering of Text) that uses NLP to automatically extract and use it to strategize and make informed decisions.

International Data Corp. has projected that productivity improvement valued at more than $400 billion could be achieved in 2020 alone as a result of knowledge gained from structured and unstructured data. The more paper the government produces, the more incentive it has to move away from producing more than it must. To that end, it has digitized roughly 250 million pages of records, with a goal of one-half billion pages by 2024. All of those mountains of data need to be structured, which is the role of NLP. Through entity recognition, unstructured data such as tweets regarding the 2020 Presidential Election can be categorized as supporting either of the major candidates to discern public opinion from raw data instead of from surveys that some think are weighted one way or another.



Source: iMerit

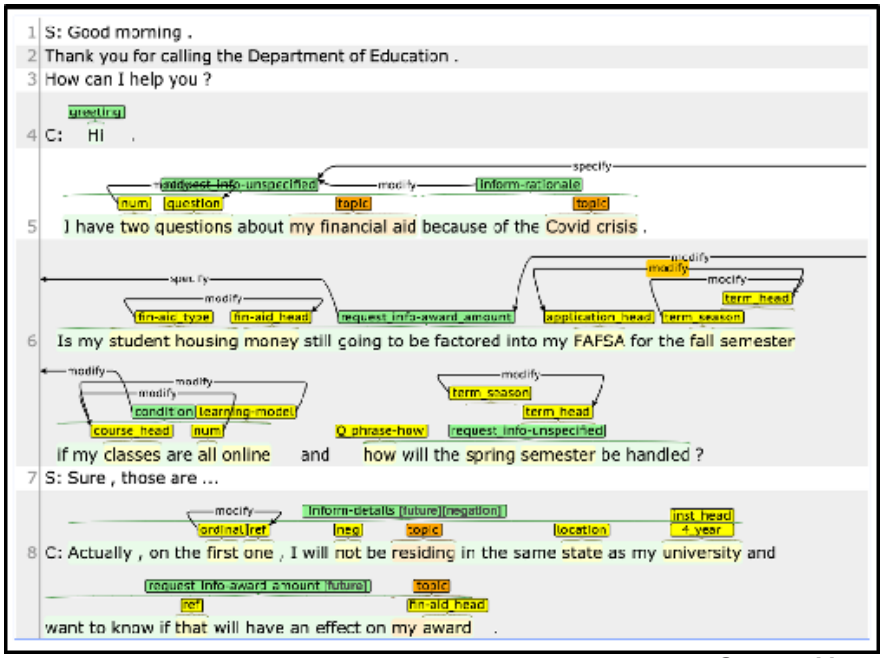
A related application is commented on by Lu Wei, a professor at Singapore University of Technology and Design, who told GovInsider, "You can build a system to detect fake news in tweets." In an age where threats of international terrorism has gone digital -- with cyber attacks on the rise -- NLP is being used by governments around the world to anticipate and detect warning signs of impending attacks.

Machines are being trained to flag semantic information from documents that can lead to identifying trouble spots or aberrations or coded language in what is otherwise normal-looking communications. Using entity classification (also known as text clustering), the U.S. Department of Agriculture's Center for Tobacco Products sorts documents concerning menthol cigarettes in one grouping to better understand and address through policymaking the smoking habits of underage youth who favor that taste.

NLP also is highly useful for analyzing patterns to determine the reasons or factors for why certain government documents are classified while others are not. Taking it a step further, NLP can be used to classify documents directly.

As paramount as public safety is for governments to monitor and keep secure, there are also a great deal of applications for NLP in the public sector that can improve everyday qualify of life. In healthcare, a study of urgent care cases can help improve the practice of triage by ranking the most serious and frequent cases that should command immediate attention.

Machine Learning for NLP is a never-ending affair, of course, because all languages evolve in terms of their common usage, with slang expressions and other departures from formal language constantly entering the vernacular. For example, millennials have their own vocabulary that older persons don't necessarily recognize. NLP has to keep up with the times, which is just one reason why unstructured text must be expertly annotated and extracted for it to do its job with almost 100% accuracy.



Source: iMerit

CV FACILITATES CRIMINAL INVESTIGATIONS

In the movie "Patriots Day," about the 2013 terrorist bombing at the Boston Marathon, officials from federal, state and local law enforcement agencies are seen in a makeshift emergency outpost endlessly studying and manipulating images playing across laptop screens. They are trying to reconstruct and make sense of the events that took place before, during, and after the deadly explosion that killed three and injured more than 250.

The so-called Computer Vision footage at their disposal was compiled from an array of capture devices. Even in 2013, video "eyes" were becoming ubiquitous on streetscapes: cell phones, surveillance cameras, broadcast footage, sensors, body cameras. They all contributed to the volume of visual, text and other data -- 10 terabytes' worth -- that led to the capture and conviction of the two suspects.

The deployment of Machine Learning to trace their activity and home in on their ultimate destinations made possible a highly accelerated process for a complex and granular investigation. One estimate is that it would have taken a human five years to manually review the 10TB of video and other data at the investigators' disposal. Computer Vision's role in the Boston Marathon case stands as a landmark, seminal application of video analytics to solve a crime.

Computer Vision uses Artificial Intelligence to study, process, and interpret both moving and still images. Government units in law enforcement, public safety, and other disciplines take advantage of the technology to facilitate criminal investigations, infrastructure surveillance, and high-level event security, where terrorism vigilance always is a top priority. Since that milestone use of visual forensics in 2013, video images have become an intrinsic part of society's central nervous system.

We cannot go anywhere these days without moving and still images filling our field of vision. Google Earth alone is proof of that. Anybody can go online and home in on any terrestrial location anywhere in the world. The average person, whether aware of it or not, is starring in his or her own movie, every day, everywhere. We constantly are under surveillance.

On top of the video devices that were exploited by the Boston Marathon data scientists, today we are virtually swimming in a rising tide of images generated by satellites, drones, dashcams, body cameras, and webcams.

Computer Vision for government use is becoming more affordable, thanks to low-cost computing and faster processing speeds. Using high-resolution images, Artificial Intelligence, and neural networks, Computer Vision can accurately mimic the way humans see, but with a substantially lower error rate of 3%, compared with 5% for humans who are analyzing images.

Market intelligence firms estimate Computer Vision will become a $17-18 billion business by 2030.

Another emerging area of CV is image recognition. In local policing activities or in vehicular tracking, it can read and record license plates. In healthcare, image recognition can more precisely diagnose illnesses. When paired with Artificial Intelligence, the technology can compare images found on social media to stored data sets to identify a suspect, and even can analyze facial expressions to extract sentiment behind a point of view. For example, the use of traffic cameras have become standard operating practice throughout the country. Law enforcement uses them for analysis of moving violations and traffic flow management, to minimize collisions. State highway authorities use overhead cameras at toll plazas for a bird's eye view of scofflaws. Video "catches" them in the act, and data science takes over to track them down and bill them for the toll, under penalty of license suspension, or revocation for non-payment.

When algorithms are optimized through Machine Learning to analyze video, it can reduce to a matter of days, if not hours, the time required to index the visual data, sorting through classifications such as a person's age and gender, an object's size, shape, or color, and a vehicle's speed and direction. For example, the Machine Learning algorithm can be trained to search for all males with a backpack from 20-25 years old, or for all cars colored green that are headed south. The power of Video Analytics continually grows more sophisticated as data scientists master how to make the most of Machine Learning.

Beyond studying recorded images after the fact, as in the Boston Marathon attack, CV holds the potential for investigators to monitor real-time activity via video that serves preventive purposes by alerting authorities to pending trouble, so they can stop it before it causes severe damage or costs lives. Tests like that have been performed in the past at airports in London, where suspicious activity was followed and flagged. The machine is trained with examples of typical human behavior so that any variations quickly can be isolated and addressed as appropriate or inappropriate. There are advances as well that detect micro-movements in a person's behavior, such as dropping a suspicious-looking bag in an airport.

The benefits for federal agencies are varied and continually expanding. They are becoming more knowledgeable and comfortable with the technology. It is being integrated as a core resource into their arsenal of assets, helping to boost the success rate of their desirable outcomes.

The U.S. Department of Homeland Safety (DHS) makes extensive use of Computer Vision to more speedily and thoroughly monitor live streams of video that can detect suspicious activity in advance of an event that can threaten public safety and people's lives. Such forms of predictive situational awareness are considered a major future use case for CV.

Agencies can invest in CV futures, as it were, by annotating and indexing archival video data, making it highly searchable. They then can use it as a control group to identify and track behavior or objects that vary from the normal values that the machine has been taught. In the same way, CV is useful for isolating fake or counterfeit images from authentic ones.

The Department of Defense's Intelligence Advanced Research Projects Activity (IARPA) unit functions as the research arm of the U.S. intelligence community. For the past several years, it has been building ever larger datasets for the purpose of refining Computer Vision algorithms to track human activity in urban environments. By daisy-chaining video data collected from unrelated networks, the agency's success rate in identifying targets could be expected to increase significantly moving forward.

The challenge for IARPA is to train algorithms that can follow human activity which spans a disparate array of video surveillance systems in far-flung locations. To accomplish this, the agency needs a lot more data to feed the machines. This way, the broadest range of possible pathways is taken into account.

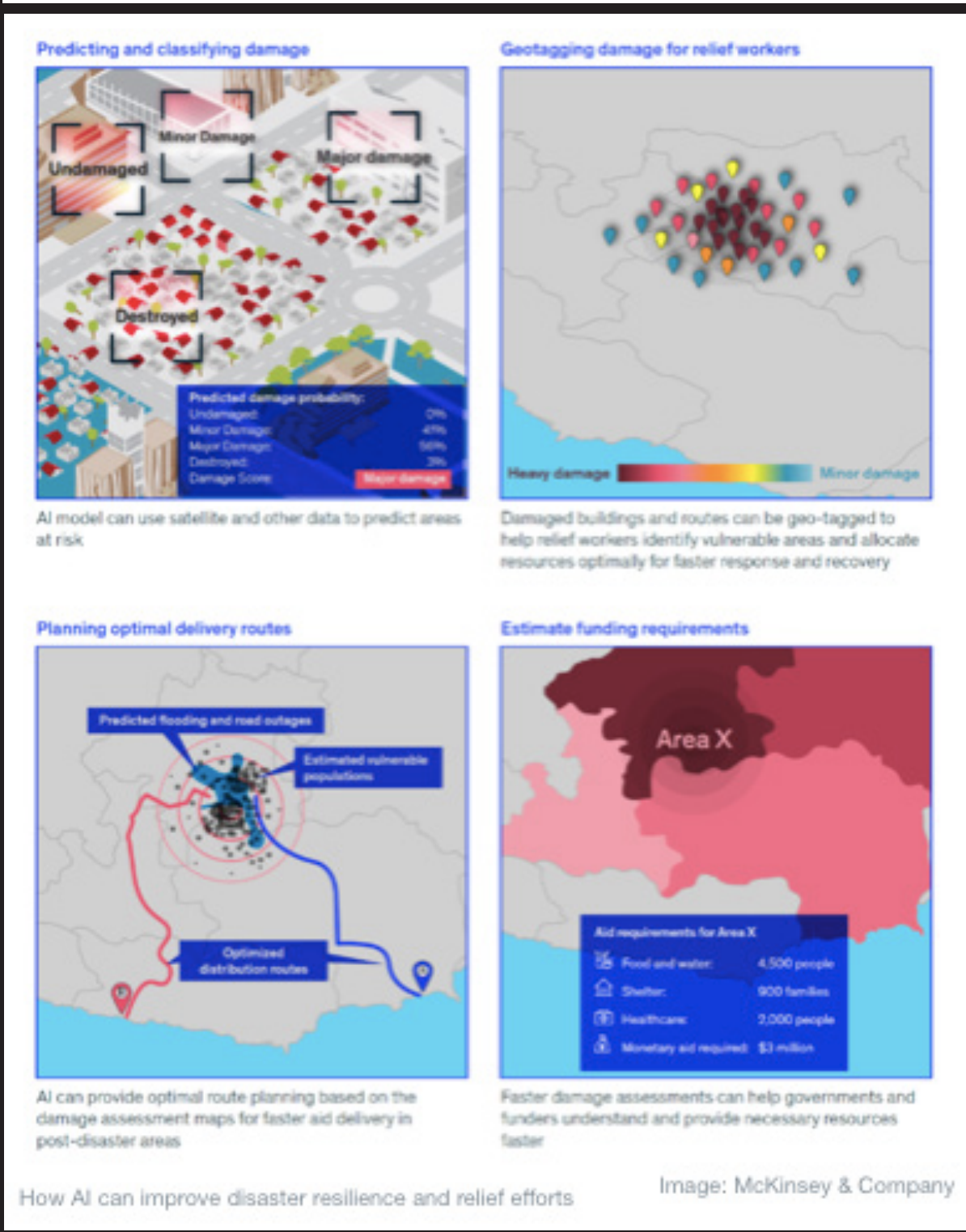
As smart as the machines are, they only know what their algorithms have been trained to recognize, and then they compare that baseline information to aberrations or anomalies for further analysis by their human operators. The implications of leveraging much larger datasets to push the capabilities of CV are profound across the public sector. Beneficiaries include the armed forces, intelligence agencies, transportation systems, environmental agencies, homeland security, aeronautics, and more.

One of the advantages of AI applications like Computer Vision is the re-allocation of a specialized, skilled workforce to focus their time and resources on high-value tasks that are beyond the scope of automation.

There are emerging a number of issues to be addressed concerning such critical matters as access to better training data and to multi-camera networks to improve Machine Learning accuracy, ethical questions of privacy, on bias (notably in facial recognition algorithms), and CV best practices that should be standardized across federal operating units to foster more cooperation and shared knowledge that will lead to more consistent quality in both video and data output.

AI FOR DISASTER RESILIENCE BY GOVERNMENTS, NGOS

Here are examples illustrating how Geospatial applications can be used to help the public sector respond to disasters. Whether natural or not, AI improves disaster resilience efforts. Regardless of continent or country, governments, NGOs and international aid workers use AI to identify damage, deliver aid fast, and rebuild communities after disasters, putting affected people on paths to recovery.



How AI can improve disaster resilience and relief efforts

Image: McKinsey & Company

THE RISE OF THE CHIEF DATA OFFICER IN GOVERNMENT

There is a visible, and significant, development taking place in government, at all levels, that underscores the bureaucracy's growing recognition of the virtually infinite value locked inside big data: the rise of the Chief Data Officer.

With the growing sophistication of Machine Learning, and its spreading utility in the public sector -- notably through Natural Language Processing and Computer Vision -- it's no longer adequate to treat data science like a nice afterthought or optional upgrade to legacy systems. Nor is it defensible for government operating units to think of it primarily as a future technology, claiming that it is premature to commit dedicated resources, both human and financial.

Silicon Valley's relentless pursuit of AI innovation and its attendant market domination of our culture should be the only motivation government operatives need to equip our bureaucracy with state-of-the-art tools to keep pace with its citizenry in terms of technological adoption. The management of data has become so pervasive and endemic to how government goes about its business that it demands its own vertical job description. And so was conceived the role of Chief Data Officer. The CDO is a logical extension of the same lineage as CIO and CTO.

The newly evolving position is institutional proof of the expanding role data has in people's lives, and not always to their satisfaction. Privacy -- or, more precisely, the invasion of it through sharing or surrendering personal data in exchange for access or information or something else of value -- is destined to always be a point of reluctant compromise for the populace. To coin a phrase, we want to have our privacy and data too.

People also are more insistent on having transparency with public data, whether it concerns healthcare, unemployment benefits, census results, or election protocols. It is incumbent on our representative forms of government to instill implicit confidence in their accountability to constituents. To that end, the benefits of deeper and more accesible reservoirs of data encompass fraud detection, minimizing wasteful spending and abuse, speedier resolutions on the disposition of workforce benefits and social services, infrastructure improvements, and overall efficiency in government services, which should translate into more reliable controls on runaway taxation.

That all adds up to increasing responsibility for the CDO, who are stewards of the treasure trove of public data that stands ready to be transformed by Artificial Intelligence into real-world intelligence that affects the everyday life of every person under the government's gaze. They also are expected to find new and better ways to improve outcomes based on the data in their purview, and to make it easier for citizens to access data that serves their interests -- whether it's about education, weather, disaster relief, agriculture, ecology, national security, economic stimulus packages, criminal justice, or housing.

A movement is afoot in support of Open Data, which would make available another magnitude of knowledge that can be shared across agencies and even with the public. Currently, that's not the case, owing to privacy and incompatible IT systems in the government complex that prevents mutual or centralized access to data. According to McKinsey Global Institute, unlocking data that's now unavailable for Machine Learning could translate into a positive economic impact of $3 trillion.

In their exhaustive report on "Leveraging Data for Economic Recovery: A Roadmap for States," Katya Abazajian and Tyler Kleykamp (of the Beeck Center for Social Impact and Innovation at Georgetown University) conclude, "by using data to inform policy and improve public understanding of pathways out of poverty, states can efficiently and effectively lead the way to improve economic mobility for their residents." They invest a great deal of faith -- and responsibility -- in Chief Data Officers to lead state efforts, responding not only to Covid-19-level crises, but other "unforeseeable needs."

Compounding the challenge for CDOs and local governments are funding shortfalls and downsized teams. "Data must be treated as critical infrastructure for emergency response," the authors contend, "in order to be prepared with insights when immediate needs arise… state leaders need to adopt more robust data-driven agendas to enable this cultural shift to occur."

Another key objective of the CDO's job is to automate processes in preparing data for Machine Learning, which can take up to 80% of a data scientists' working hours. The workload involves preprocessing and normalizing data, filling in values that are missing, choosing in what format to interpret the data; identifying variables in the data; and testing thousands of algorithms.

Tools to help CDOs better fulfill their job functions are becoming more prevalent, which in turn will ease government's progressive adoption of Machine Learning. One of those tools is synthetic data, which greatly speeds the painstaking process of acquiring and training millions of data elements. In a test of the new technology by Deloitte, the actual training data required was reduced to 20% of the total data, with the remaining 80% essentially cloned. The task tested was analyzing job titles and descriptions. Several thousand examples were used in training the algorithm.

With CDOs leading the way, it is hoped that both civil servants and citizens will reap the benefits of a progressively vital government data infrastructure committed to growth in positive outcomes and accountability.

HOW NLP HELPED COMBAT COVID IN COMMUNITIES

During the Covid-19 crisis, the Kansas Department of Health and Environment used Natural Language Processing to extract data from cellphones to track citizen compliance with stay-at-home mandates. There even was a report card at the community level to grade (from A-F) how well locals obeyed social distancing and stayed away from off-limits destinations where crowds could gather in contravention of state guidelines. The data analytics bore further fruit for state authorities by pointing them to spots where non-compliance was likely to cause virus cases to spike. Similarly, law enforcement in Minnesota obtained state records documenting the location of those with Covid-19 so they could stay vigilant in protecting the greater community. The same is being done in other states, allowing EMT crews to stay aware if they might be at risk when responding to a situation where Covid-19 is a threat. That's the good news.

Within the larger context of the vast federal enterprise, however, Covid-19 also has laid bare how far the U.S. government still needs to go in making optimal use of NLP, whether it's at Health and Human Services (HHS) or the Veterans Administration (VA) or NASA, or any other agency. There are multiple challenges, with middle managers unskilled (or uninterested) in how to manage NLP and ML; inadequate budgets to acquire the capital equipment necessary for data analytics; a scarcity of datasets; no standards for how to gather and handle the large datasets required for well-informed outcomes. As a result, "Covid-19 has established how poor our data infrastructure is," says Dr. Suman De, a government specialist at Infosys Public Services, in a report on AI's potential role at Health and Human Services by Deloitte.

The Deloitte report concludes the pandemic calls for the kind of "accurate, real-time data and analytics" required "not only to track the spread of the virus but also to limit its impact through proactive case management, and to make informed decisions about stay-at-home orders and other restrictions." Dr. De adds that agencies "are not able to act on information" due to "a lack of proficiency in analytics that is necessary to improve outcomes." In some areas of the United States, people can check for Covid-19 symptoms by using chatbots, which supplement health workers and ease their burden.

CHATTING WITH ROBOTS IN THE PUBLIC SECTOR

Chatbots are a highly visible and rapidly expanding manifestastion of Natural Language Processing (NLP). We've all either used them or been invited to engage with them when they pop up in the corner of a website we're visiting. They offer the visitor instant help without having to place a call or send an email. Currently, it's believed up to 25% of organizations employ a virtual customer assistant, says Gartner. It's forecast that by 2023, 25% of the U.S. population will use digital assistants on a regular basis. Even today, according to a GSA estimate, the federal government fields three-quarters of a billion public inquiries a year using chatbots. Its budget to handle those calls through contact centers is $17.5 billion.

The role of Natural Language Processing is to match user text or voice input to executable commands. Digital assistants with voice recognition turn spoken language into text. By training large datasets, conversations can be personalized to the point an end-user may believe they are communicating with a human on the other end. A common thread in the ongoing dialogue on Artificial Intelligence in government circles is the democratization of the technology. In this context, it means putting the tools, techniques and training regimens for such AI applications as Machine Learning and NLP within reach of a wide swath of federal or local government analysts and engineers to encourage best practices and easily shareable data across municipalities and agencies.

The far-reaching goal is for ML and its applications to become as useful and accessible a tool as a cubicle workstation (as evidenced by the Deloitte statistics cited earlier of more than 80% public sector companies wanting to bring AI into their enterprise). At the state and local level, it's a different story. According to a survey by the Center for Digital Government, while half of leading companies in the private sector use Artificial Intelligence, less than 16% of state and local officials use the technology, and less than 30% intend to adopt it any time soon.

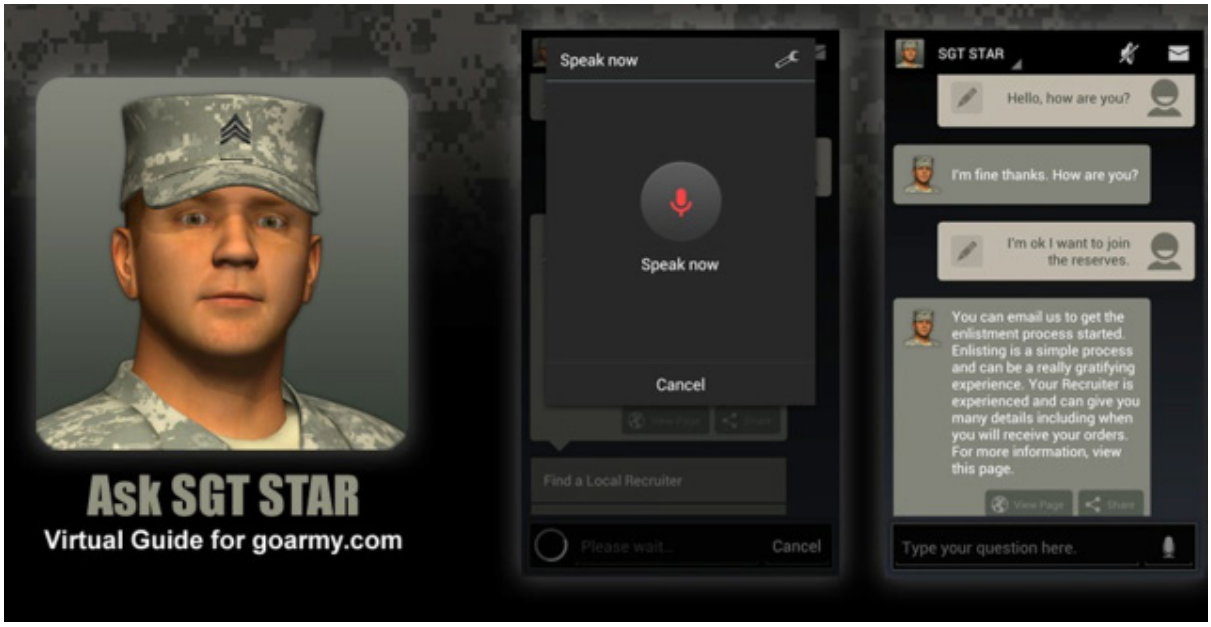
But the one area where governments at all levels are warming to AI is for the more familiar form of democratization: keeping citizens informed about their rights, about civic services available, and about how to navigate bureaucratic red tape. There is no more elementary or commonplace form of NLP than chatbots, and the voice assistants we know as Siri and Alexa, found in smartphones, smart speakers, TVs, websites, and a growing number of digital devices. As Utah Chief Technology Officer Dave Fletcher says, "Technology vendors are embedding AI into everything."

The word chatbot is a portmanteau (two words combined into one) that merges chat with robot. It can carry on a conversation with a human. Its knowledge base, which informs the conversation, is determined by the richness of the dataset that has been trained, through ML, to operate the chatbot. What has accelerated the spread of chatbots is their relative simplicity as AI tools, plus the maturity of ML, allowing algorithms to advance in sophistication, thanks to years of data training.

They not only can process voice recognition into text with exacting accuracy, but

also can detect the sentiment in the user's voice commands or questions. The NLP technology is adaptable enough to even understand slang expressions whose meaning may need to be analyzed, interpreted and, in effect, decoded by the machine. For heavily-trafficked government sites, like the Social Security Administration, automating the interaction with tens of millions of "customers" frees the time and focus of agency personnel to attend to other tasks that are not as easily or effectively or economically automated. The Singapore Government has been using the "Ask Jamie" chatbot across all agency websites for the past few years. In 2016, China's largest search engine, Baidu, launched a chatbot that assists doctors in answering patients' questions and suggesting treatment options.

For chatbots deployed in the service of citizens to be useful and effective, a sizable volume of data is essential. That either can be in the form of government-sourced information, such as recycle pickup schedules and parks and recreation activities, or it can be open data, which is equivalent to public domain content: it is not protected as intellectual property legally owned by someone, so anyone can use it or redistribute or share it. Chatbots are a logical high-tech replacement for 311 call centers administered by state or local governments, which can entail costly and labor-intensive phone banks requiring a sizable workforce to meet the expectations of constituents.
.



At the federal level, the earliest chatbot on record is Sgt. Star, a digital doppelganger "enlisted" by the U.S. Army in 2006 to engage with recruit prospects by answering their questions. Modeled after an actual soldier, the virtual human was created by the Institute of Creative Technology at USC. The efficacy of using a chatbot like Sgt. Star is evident in the metrics. Recruits on average spend about 10 minutes with Sgt. Star, compared with less than half the time spent before he was put into action. The Army reports he has proved equivalent to 55 human recruiters in his

workload, and has addressed in excess of 16 million questions, with an accuracy rate of 94%. Emma is the name U.S. Citizenship and Immigrations Services has given to its virtual assistant. There have been millions of questions Emma has fielded, from English and non-English speaking visitors. A spokesperson for the agency acknowledges the role of Machine Learning by pointing out that every time a new question is put to Emma, it helps the agency's data scientists improve her future responses.

At the U.S. Department of Agriculture (USDA), research showed that not only farmers but its own workforce found it too inefficient and inconvenient to search for answers to their questions at USDA.gov. As a result, the agency decided to develop a chatbot from the ground up that would improve the user experience for both customers and staff. Delving into its vast Farm Service Agency (ASK FSA) knowledge base, USDA identified 75 categories about which most questions were asked. The chatbot design team restructured the knowledge articles to conform to how users queried a virtual assistant. The USDA chatbot has yet to go live, but the agency says it is committed to using Artificial Intelligence and Natural Language Processing more extensively, with an eye toward upgrading its customer service in responding to queries, as well as to better manage costs as it scales AI deployment, which it expects is true of other federal agencies.

By relying more on AI for data analysis and customer-facing services, it looks forward to re-allocating human resources to focus more quality time on high-touch tasks that are best managed by human intellect and analysis. In that way, government efficiency is optimized as agencies are able to operate more quickly and responsively to public needs.

Virtual assistants also are used internally to streamline workflow. Agencies such as Treasury, NASA, GSA (General Services Administration), and VA (Veterans Affairs) use bots to sort mail, organize archival materials, filling forms with boilerplate information, and other mechanical tasks that can be automated.

iMerit

GOVERNING
WITH (ARTIFICIAL) INTELLIGENCE

imerit.net