RESPONSIBLE AI:
A Framework for Building Trust in Your AI Solutions
Artificial intelligence (AI) is maturing rapidly as an incredibly powerful technology with seemingly limitless application. It has demonstrated its ability to automate routine tasks—such as our daily commute—while also augmenting human capacity with new insight. Combining human creativity and ingenuity with the scalability of machine learning is advancing our knowledge base and understanding at a remarkable pace.

However, with great power comes great responsibility. Specifically, AI raises concerns on many fronts due to its potentially disruptive impact. These fears include workforce displacement, loss of privacy, potential biases in decision-making and lack of control over automated systems and robots. While these issues are significant, they are also addressable with the right planning, oversight, and governance.
Responsible AI is a framework for bringing many of these critical practices together. It focuses on ensuring the ethical, transparent and accountable use of AI technologies in a manner consistent with user expectations, organizational values and societal laws and norms. Responsible AI can guard against the use of biased data or algorithms, ensure that automated decisions are justified and explainable, and help maintain user trust and individual privacy. By providing clear rules of engagement, responsible AI allows organizations under public and congressional scrutiny to innovate and realize the transformative potential of AI that is both compelling and accountable.

This is the crux: Federal agency leaders are increasingly being asked to answer three critical questions regarding their planned use of AI:

01 How do I ensure that AI solutions are designed to operate responsibly?

02 How do I employ AI to act in a manner that is compliant with stakeholder expectations and applicable laws?

03 How do I use AI to unlock the full potential of my workforce?

Given the significance of their decisions and the scrutiny that they operate under, Responsible AI should be a priority for federal agencies. Furthermore, close adherence to Responsible AI principles can allow federal agencies to set broader expectations for the positive use of AI to benefit society.
We already see the transformative power of AI technologies in the commercial and public sectors, with autonomous cars, chatbots that converse using natural language, and warehouses staffed by robots tirelessly picking and packing day and night. We have also witnessed the downside of this transformative technology, with autonomous car accidents, chatbots that learn and mimic offensive language, and the threat of displaced workers. These incidents have stoked fears of a “Job Apocalypse,” or super-intelligent AI, as well as specific concerns regarding inclusion, diversity, privacy, and security.

As AI technologies become more pervasive, deeply embedded in existing solutions and responsible for an increasing number of decisions like benefit payments, mortgage approvals, and medical diagnosis, they become less visible and transparent. Unlike the autonomous car or the warehouse robot, algorithms are not visible. And organizations face ethical as well as legal and regulatory risks if they use a “black box” approach to AI.

One of the real risks with AI is amplifying and reinforcing existing human biases. Some of these biases are unintended and come about due to a lack of diverse perspectives when developing and training the system. In other examples, decision-making can be skewed by reliance on incomplete data where other relevant factors are omitted. Finally, historical data, on which the system will train, may be incompatible with modern attitudes, especially around individuals’ roles in society as it relates to gender, race and similar attributes.

In addition to being free of bias, AI-based decisions should be understandable to those impacted and adhere to existing rules and regulations. For example, the 1974 Equal Credit Opportunity Act has long required that those denied credit must be advised as to the reasons behind that decision. More recently, 2011 Federal Reserve System SR 11-7: Guidance on Model Risk Management
advised banks to employ active risk management to guard against faulty assessment models. In New York City, the Council recently created a task force charged with determining which “automated decision systems” used by the city should be subject to further governance procedures that include: 1. allowing citizens to request an explanation as to how decisions were derived using these systems, and 2. whether those decisions disproportionately impact age, race, creed, color, religion, national origin, gender, disability, marital status, partnership status, caregiver status, sexual orientation, alienage or citizenship status.

In a similar vein, steps should also be taken to ensure that AI doesn’t inadvertently uncover Personally Identifiable Information (PII) and similar data in supposedly anonymized data. This will place greater pressure on data supply chains and the need to ensure that data is used consistently with the expectations that it was originally given. Beyond traditional data sources, the more widespread use of computer vision for observational monitoring and analysis opens new questions for society at large.

According to Accenture Research, 82% of federal executives believe that “AI will work next to humans as a co-worker, collaboration and trusted advisor” within 2+ years. A high degree of trust will be required for the workforce to increase their reliance on automated systems for often life-impacting decisions. This trust can develop from a widespread understanding for how these decisions are made, ability to guide the machine as it learns, as well as knowledge about how humans and machines augment each other for improved outcomes.

In addressing all these issues, Responsible AI offers a way for all stakeholders to adopt a ‘people first’ approach that is fair, accountable, honest, transparent and human-centric.
Responsible AI Primer

Responsible AI is about creating governance frameworks to evaluate, deploy and monitor AI to create new opportunities for better citizen and mission services. It means architecting and implementing solutions that put people at the center. By using design-led thinking, organizations examine core ethical questions in context, evaluate the adequacy of policies and programs, and create a set of value-driven requirements governing AI solutions.

The four foundational elements of Responsible AI include:

**GOVERN** – You must create the right framework to enable AI to flourish—one that is anchored to your organization’s core values, ethical guardrails, and regulatory constraints. Standards bodies such as IEEE are providing guidance for global organizations to ensure every stakeholder involved in the design and development of autonomous and intelligent systems is educated, trained, and empowered to prioritize ethical considerations.¹

**DESIGN** – Any new solution should be architected and deployed with trust built into the design. This means that requirements for privacy, transparency, and security have equal weight with new product features. The resulting systems should address the need to include AI solutions that can explain their rationale for decisions making. Capital One is researching ways to make AI more explainable, hoping to use it to review credit card applications since banking regulations require that financial companies furnish an explanation to customers when their applications are denied.²

**MONITOR** – AI needs close supervision using ongoing human monitoring and auditing of the performance of algorithms against key value-driven metrics such as accountability, bias, and cybersecurity. Automakers Volvo³ and Audi⁴ are addressing accountability with announcements that they will assume liability for any accidents that happen when automated driving technology is in use.
Cases of bias may be subtle and more difficult to catch and thus require close attention. Accenture is developing a tool to help businesses detect gender, racial and ethnic bias in artificial intelligence software. It lets users define the data fields they consider sensitive—such as race, gender or age—and then see the extent to which these factors are correlated with other data fields. Most importantly, it creates the feedback needed so that teams can govern AI and make adjustments to address bias.

**TRAIN** – According to the forthcoming Accenture report—AI and the Federal Workforce—74% of federal workers believe that it will be somewhat, very, or extremely important for them to develop skills to work with AI. Therefore, it is incumbent on agencies to prepare and equip them to take full advantage of AI and the new workstyles that it fosters.

This requires training in the near-term to better understand how AI systems operate and potential upskilling long-term. As federal agencies will need to prepare their workforces to work with AI, an integrated approach is needed, including:

- **Educate employees to how AI will be integrated into operations and why**
- **Ask employees where and how AI might improve their day-to-day roles**
- **Engage employees in co-creation to determine how people, processes and AI technology come together to create a more intelligent agency**
- **Develop the skills needed for employees to take advantage of the insight offered by AI to achieve better, more consistent outcomes**

With technology-led changes accelerating throughout the economy, investing in continuous learning to maintain a qualified workforce must be a priority.
What is Explainable AI (XAI)?

Explainable AI (XAI) is defined as systems with the ability to explain their rationale for decisions, characterize the strengths and weaknesses of their decision-making process, and convey an understanding of how they will behave in the future.

NVIDIA is one company that is tackling the black box issue head-on. Drive PX, NVIDIA’s AI-infused self-driving car platform, was designed to “teach” itself to drive but the way it did so was not entirely clear. To improve the system, NVIDIA engineers prioritized opening the AI black box and developed a way to get a Drive PX vehicle to explain its driving style visually. The platform does so by displaying a video of a recently driven streetscape, over which it then highlights areas that it gave the most weight to during navigation.

An open issue for XAI is who are the explanations for—advanced mathematicians or engineers, or employees and customers? Also, much of the AI employed today automates more traditional statistical methods, which are more easily explained than neural-net-based decisions used for image recognition or self-driving cars. Accenture is forecasting the emergence of new “explainer” roles to help fill this void, but work remains to make AI models more readily interpretable.

There are two notable efforts to create XAI: DARPA-XAI and LIME (Local Interpretable Model-Agnostic Explanations). US Department of Defense’s Defense Advanced Research Projects Agency (DARPA) launched the Explainable Artificial Intelligence (XAI) program to identify approaches that will give AI systems “…the ability to explain their rationale, characterize their strengths and weaknesses, and convey an understanding of how they will behave in the future.” LIME is a technique developed at the University of Washington that helps explains predictions in an “interpretable and faithful manner.”
As you evaluate new AI solutions, there are two critical areas of impact to consider: designing solutions imbued with core values and transforming your workforce to ease the transition to new ways of working. Trust and transparency are the core principles and self-reinforcing mechanisms that ensure AI innovation occurs within the context of accountability, security, and fairness.

**DESIGN SOLUTIONS WITH AI**
Designing solutions using Responsible AI frameworks covers four interrelated areas: human-centered design, governance, training data, and monitoring performance.

**Human Centered Design**
Great design begins with empathy—having a deep understanding of user needs, and human-centered design is ideally suited to uncovering latent, unexpressed needs. The design process in this context needs to draw out user concerns about privacy, security, and trust to inform the design of the solution. This is critical given the complexity and often lack of transparency around AI solutions and how they work.

A recent study by MIT on genetic privacy policies point to potential best practices for handling privacy issues in the future. Patients asked to participate in a study were given two options; provide access to their genetic data and have control over this data or be warned about privacy risks and asked to consent to those risks with no further control over their data. For the first group, the incidence of testing increased by 83 percent. The second group saw a decrease of 69 percent.⁶
**Governance**
Implementing governance for Responsible AI requires addressing organizational leadership, project teams, and customers. Leadership needs to develop the communication channels, cultivate a culture of responsibility, and create internal governance processes aligned with regulations and industry best practices. Commensurate with this, project teams need to embed transparency into their daily practices and processes while implementing governance mandates developed by the leadership team. The focus for customers should be to engage and educate, providing a consistent level of transparency based on the principles of Explainable AI.

**Training Data**
AI uses machine learning algorithms and real-time data to improve continuously. However, the self-correcting aspect of this process should not be left solely to the machines. Self-changing, data-driven processes need human workers who can act rapidly on the opportunities that machines discover in real time. They also require people to continuously assess the need for improvements to safety, fairness, and auditability by training the algorithms over time. Accenture believes this will create a new category of job in the “missing middle,” where humans help machines and machines help humans to achieve better outcomes collectively.

**Monitoring Performance**
Model monitoring ensures AI models continue to function as intended after deployment into production. Model monitoring should include the following four processes: 1. active monitoring,
2. performance drift, 3. operational bias review, and 4. model retraining. Active monitoring tracks user behavior to identify irregular patterns that may indicate signs of unintended consequences. Performance drift monitors model KPIs to determine model performance and triggers a retraining process. Operational bias review tracks model inputs and outputs to identify irregularities which may indicate bias. Model retraining uses new data to account for changes in user behavior. By using a combination of these methods, you can more readily spot and correct irregularities or bias in your results.

WORKFORCE WITH AI
Agencies need to build workforce trust in the smart machines that employees will increasingly rely upon. They can do so by teaching them how to interact, train and augment these systems. Organizations that fail to take these steps will find many of the benefits of AI elusive and may encounter a talent crunch within the next few years.

Over the long-term, active upskilling may be required for many roles. Through education and training, valued employees can move from executing rote tasks to providing more in-depth analysis requiring greater judgment. Beyond benefits for the government workforce, federal agencies can help establish the best practices needed for a transitioning economy. As Reps. Will Hurd (R-TX) and Robin Kelly (D-IL), Chair and Ranking Member of the Oversight & Government Reform Subcommittee on Information Technology, wrote in a recent congressional whitepaper, Rise of the Machines, “the federal government should also lead by example by investing more in education and training programs that would allow for its current and future workforce to gain the necessary AI skills.”
Conclusion

Simply put, AI represents a new way of working. It will bring about profound changes within organizations and society that we can’t fully understand and predict today. In this context, responsible AI is a critical component of an organizational change model that focuses on rapid learning and adapting. Accenture’s AI-driven organizational change model provides a framework for how agencies can begin this process. It starts with defining your AI journey and strategy, followed by building a talent strategy and learning architecture. Finally, organizations execute rapid upskilling programs and closely measure progress.

Your AI strategy should include an AI operating model with identified stakeholders and objectives, a defined AI talent strategy, required ecosystem partners, and initial prototype initiatives. To build a talent strategy and learning architecture you need to design AI-specific talent profiles and segments with competency frameworks, develop the required competencies, and design curriculum and learning paths. For rapid upskilling, agencies need to build and launch targeted learning initiatives and monitoring progress against established goals.

Organizations that want to maintain momentum for AI-driven change should consider creating specific AI Centers of Excellence (CoE) that provide advisory support, engineering and development, research and learning, communities of interest across the enterprise, and governance and change management. This is a model that the U.S. Department of Homeland Security recently implemented. CoEs are an essential mechanism to ensure responsible AI practices are followed, and solutions are developed in a responsible manner.

By embedding responsible AI into your approach for organizational change, you ensure that the critical element of trust is cultivated and maintained among key stakeholders, the most important of which being employees and customers.
ABOUT ACCENTURE FEDERAL SERVICES

Accenture Federal Services, a wholly owned subsidiary of Accenture LLP, is a U.S. company with offices in Arlington, Virginia. Accenture’s federal business has served every cabinet-level department and 30 of the largest federal organizations. Accenture Federal Services transforms bold ideas into breakthrough outcomes for clients at defense, intelligence, public safety, civilian and military health organizations. Learn more at www.accenturefederal.com.

ABOUT ACCENTURE

Accenture is a leading global professional services company, providing a broad range of services and solutions in strategy, consulting, digital, technology and operations. Combining unmatched experience and specialized skills across more than 40 industries and all business functions—underpinned by the world’s largest delivery network—Accenture works at the intersection of business and technology to help clients improve their performance and create sustainable value for their stakeholders. With 459,000 people serving clients in more than 120 countries, Accenture drives innovation to improve the way the world works and lives. Visit us at www.accenture.com.

REFERENCES

1https://standards.ieee.org/industry-connections/ec/autonomous-systems.html

Copyright © 2018 Accenture
All rights reserved.
Accenture and its logo are trademarks of Accenture.