

DATA STRATEGIES FOR AN AI-POWERED GOVERNMENT

Key findings and recommendations for the federal enterprise from planning to piloting to procurement.

The rapid increase in demand for the use of artificial intelligence (AI)-based tools through cloud service providers poses significant challenges for the data ecosystem within the federal information technology (IT) enterprise. The Atlantic Council GeoTech Center examined what federal agencies can do to maximize the value of data through AI and establish an AI-ready data infrastructure, and identified strategies, tactics, approaches, best practices, and lessons learned to accomplish these goals. Four key focus areas were identified.

1. Establishing human capital and an “AI-ready” culture

Key findings:

- Addressing human capital and workforce challenges is crucial when dealing with AI and data management. It is critical to integrate humans into the AI and data management process and obtain leadership buy-in on approaches that balance other concerns, such as security.
- There is a broad need to improve AI literacy across the enterprise, especially at the leadership level, and to develop more acceptance of necessary risk related to AI applications.
- Users will find value in supporting tools that augment their capabilities but do not take over their decision-making.
- For organizations that have not routinely leveraged data for analysis or policy insights, identifying and socializing mission-specific needs and insights that can be addressed helps establish an initial stakeholder community.

Agencies should consider:

- Creating *cross-functional* task forces and working groups to get data AI-ready and identify key questions where data and AI might reveal actionable insights, promoting organizational adaptation concurrent with technological change.
- Rewarding risk-taking on AI projects, balancing “misuse versus missed use,” and encouraging an approach of “yes, unless” for data sharing.
- Examining roadblocks to moving each use case forward and ensuring the necessary workforce is available given the scale of each problem.
- Sending clear demand signals and explaining the value proposition and scalability of data-centric AI applications, making clear the return on investment and measures of effectiveness.
- Working with service providers to understand how they use AI and how to use AI through their services.

2. Planning and developing data-centric AI applications

Key findings:

- Agencies have an overwhelming quantity of data that creates unique data governance challenges.
- Data is often poorly structured and siloed in ways that hamper the efficient use of information.
- AI relies on data, but leaders need to recognize it must be applied to the right data and problem.
- It is important to consider the value of specialized data used for specific applications versus more generalized data applied to broader solutions.
- AI-generated synthetic data can be used to protect privacy with the caveat that its recursive use can lead to spurious output.
- It is still a challenge in the health sector to convert images, such as faxes, into structured data.
- The White House is working to help agencies turn data into action by collecting data purposefully in such a way that they can more easily parse it and achieve equitable outcomes.

Agencies should consider:

- Operationalizing data repositories into a data fabric with organization-wide access.
- Establishing a dedicated point of contact for data repository requests.
- Ensuring that customers know where their data is and who owns it.
- Treating data as a product that requires trust and continually seeking feedback on its use.
- Evaluating what applications can be developed with existing data before collecting new data and proactively considering what applications might arise from new data before collecting that data.
- Working across the interagency to create common tags for fair data and shared test data sets.
- Integrating privacy principles from the start, using appropriate types of encryption, where required, along with appropriate access controls.
- Stratifying applications based on risk.

3. Piloting data-centric AI applications

Key findings:

- Managing and maintaining the data pipeline is key, from getting the data to cleaning and organizing the data to deploying the data. However, it is difficult to get program managers to think strategically about data up front, resulting in myriad challenges later.
- When it comes to more specialized data, one must prioritize quality over quantity. There may be pressure to “go big” or “not at all.”
- During pilots it is important to integrate the application with human systems, getting it into the hands of users and continuously obtaining feedback, reexamining the data, and updating the software in real time.

Agencies should consider:

- Embedding the technology with operational users as quickly as possible for immediate feedback and to identify unanticipated problems through extensive testing, including infrastructure and data challenges.
- Being flexible, agile, and forward leaning with people on the “forward tip of the spear” and embedding data professionals in projects who understand both the data and the mission.
- Picking key anchor projects that have high leverage potential.
- Identifying applications and pilots that could be expanded across sectors/organizations.

4. Procuring and/or scaling data-centric AI applications

Key findings:

- Commercial providers require data of sufficient quality to engender trust in the resulting AI applications. Partnerships with the private sector are needed to move the needle across the US government.
- A promising area to scale is leveraging large language models’ (LLMs’) ability to write code and find bugs as effectively as “bespoke” application development tools.
- A challenge to scaling LLMs/generative AI at this time is that hallucination rates can approach 30 percent, too high for widespread use.
- Generative AI introduces new threats that must be acknowledged and rapidly addressed, including in the physical world where agents could be capable of autonomously causing harm.

Agencies should consider:

- Involving downstream users from the start of the transition process and getting their feedback.
- Developing a culture and narrative that makes clear that the agency is not deploying AI without considering broader applications and future scale.
- Identifying solutions in the start-up community that can be shaped for different applications along with emerging capabilities that could be useful soon.
- Moving to contract using industry best-of-breed design principles and flexible acquisition authorities when available.
- Building transparency, testing and evaluation, privacy safeguards, and other elements of responsible AI into planning and procurement.

Acknowledgments

These findings and recommendations were produced by the Atlantic Council GeoTech Center following private discussions with IT, data science, and AI leaders and experts in both the public and private sectors. This effort has been made possible through the generous support of Accenture Federal Services and Amazon Web Services.

**Accenture
Federal
Services**

powered
by **aws**

 **Atlantic Council**
GEOTECH CENTER