DATA— Fueling Artificial Intelligence

Data management must address all phases of data lifecycle, and each phase involves people, processes, technology, and governance.

We live in an age when technology empowers government, and artificial intelligence (AI) is enhancing, accelerating, and automating key decision-making for federal agencies.

AI can streamline citizen inquires and processing of transactions while improving the user experience, as well as enabling government employees to spend their time on higher-value activities. And the good news is with every passing day, AI is easier to implement and use.

The Data Challenge

While AI might be at our fingertips and agencies are quickly establishing AI initiatives, AI models are very dependent on the quality of the datasets used to develop, train, and drive the algorithms. If the data is inaccurate, has built in biases, is not complete, has inconsistent data formats and metadata, or is not relevant to the business problem, it can negatively impact the integrity and fairness of the algorithm results.

Joseph Klimavicz
CIO, Department of Justice (retired)
Managing Director
KPMG LLP

AI algorithms also require massive quantities of data, and not all organizations have access to massive data. This requirement can be exacerbated when training and testing must occur with unclassified data and then operating the AI algorithms in a classified environment with additional classified data.

Additionally, privacy regulations may limit the use of sensitive data for training models. Data issues can arise anywhere along the data lifecycle; from finding the data, getting it, preparing it, processing it, and maintaining the data.

Improving Data Quality

After your workforce, data is every agency's most important asset because data is critical to all essential services, not just AI. Key aspects of a high-performing data management program include:

- Building a data strategy with a value proposition, guiding principles, goals, and objectives
- Establishing data governance to facilitate data stewardship, establish controls, and monitor data use

ML AI

- Creating a data architecture to standardized use and optimize value of data across the enterprise
- Promoting data standards to address data quality issues as you use the data
- Developing a Data Implementation and Investment Plan to document specific actions to close the gaps
- Developing a data fluency program that includes everyone and never misses an opportunity to educate

First step in executing a successful data management program is appointing a Chief Data Officer who is resourced and accountable to lead cultural change and oversee data strategy initiatives and is capable of building alliance among business and mission leaders.

Second step is establishing a data strategy and formulating a change management/communication plan to drive lasting cultural change through effective communications and project management guidance.

Third, establish enterprise data governance that defines policies and decision-making bodies, and includes monitoring, evaluating, and reporting on data use, data quality, data reuse, and data sharing.

Data governance can proactively balance, remove, or control inconsistencies that may introduce harmful bias. Data approaches should be reviewed to assess the ethical use of AI, address bias in AI testing and training by reviewing exclusion of certain datasets or imbalanced distribution and confirm that representation of individuals or groups is based in fact.

Also include tooling to monitor how model outcomes evolve and trigger points to recalibrate and validate that only relevant and reliable data is used.

Fourth, create a Data Implementation and Investment Plan to drive and resource the Data Strategy initiatives.

> The AI model development should account for the volume, complexity, and quality of data to provide a tailored AI approach.

Lastly, establish a data architecture with data standards based on industry best practices, and initiate a data inventory, and publish implementation guidance to cover logical data reuse, services integration, and data housekeeping to support data lineage.

Robust Management Strategy

After executing the above steps, identify the organizations master reference data, and promote enterprise-wide metadata standards to enable data visibility, understandability, interoperability, and trust.

Develop a data maturity model and perform regular maturity assessments to benchmark and evolve the data strategy implementation to address AI innovations. Consider 'low code' platforms and robotic process automation to ease the implementation of your data strategy.

Finally, develop and maintain a funded data fluency program that includes teaching decision-makers about sources of data bias and implications.

In summary, having a robust data management strategy is critical for the success of AI. Nevertheless, data management remains a challenge because it is a complex process. Data management must address all phases of data lifecycle, and each phase involves people, processes, technology, and governance.

Domain knowledge, acquired over decades, should be leveraged in addressing data quality issues. The AI model development should account for the volume, complexity, and quality of data to provide a tailored AI approach.

No matter how significant the efforts to address data quality, AI algorithms will have to deal with inaccurate or incomplete data, and alert users of potential issues. ■

About The Author

Mr. Klimavicz is a Managing Director with KPMG LLP where he leads the government Chief Information Officer (CIO) advisory practice and helps government clients develop and implement digital transformations. Mr. Klimavicz's 37-year career in the federal government began with the Central Intelligence Agency (CIA) as a scientist and culminated with the U.S. Department of Justice (DOJ) as Deputy Assistant Attorney General and CIO from May 2014 until March 2020. Mr. Klimavicz also served as National Oceanic and Atmospheric Administration (NOAA) CIO and Director, High Performance Computing and Communications from 2007 until 2014, and as the National Geospatial-Intelligence Agency Deputy CIO from 2003 to 2007. In 2012, Mr. Klimavicz received the U.S. Presidential Rank Award for Distinguished Executive Service, and he is a CIO-SAGE at the Partnership for Public Service.