



Generative AI and Government Data: Balancing Innovation, Security, and Privacy

Highlights from a Roundtable hosted by the Advanced Technology Academic Research Center (ATARC) in partnership with Elastic, January 2023

Generative AI holds immense potential for transforming government operations, but agencies must balance innovation with security, privacy, and ethical considerations. In a recent roundtable discussion hosted by the Advanced Technology Academic Research Center (ATARC) in partnership with Elastic, Federal experts discussed the importance of authoritative data as it pertains to generative AI use in the Federal government.

The Importance of Data


Roundtable participants discussed the opportunities and challenges presented by generative AI and the use of government data. Generative AI has the potential to help government agencies interact more effectively with the public. For instance, instead of searching multiple websites, citizens can consult a Federal GPT and receive accurate information in seconds.

Generative AI models trained on government data can not only empower citizens, but also mitigate many risks inherent with open source models. There are many known risks associated with generative AI, including the cascading consequences of hallucinated responses. Treating government data as a commodity and prioritizing proper data management are critical to maintaining high levels of security, privacy, and ethics when it comes to generative AI use.

Ultimately, this is only possible when departments operationalize data management, which is a monumental task. Most government agencies are working to gather, standardize, and qualify vast amounts of data in order to use in LLMs. Scalable solutions to operationalize generative AI in government agencies do not exist due to data classification challenges.

Panelists emphasized the emerging challenge of maintaining data integrity in an era of increasingly sophisticated AI capabilities. As deep fakes and voice fakes become more mainstream, transparency in datasets will become increasingly more important. Generative AI is enabling everyone to produce more data, which makes it more challenging to trust the outputs.

“We’re getting into an era of integrity, not just confidentiality.”



While the future of generative AI in the Federal government is promising, it must be done responsibly and ethically. AI ethics frameworks have guided responsible AI use over the past decade, and similar frameworks could be translated to generative AI.

Shifting Security Concerns

As one panelist suggests, “security in itself is going to turn into its own threat”. While increased data accessibility offers numerous benefits to both the citizen and Federal workforce, it also raises concerns about the possibility of connecting disparate information that was once very hard to access.

This underscores the need for data classification. Generative AI will only exacerbate problems with misinformation, privacy, and security if misclassified data is training large language models. Unfortunately, agencies do not have the capacity to classify and tag the vast amounts of data they are creating.

Considerations for an AI Future

While private sector AI may surpass the government’s ability to develop cutting-edge language performance, the benefits of a future government language model lie in the quality and accuracy of the underlying data used to train its model. Agencies should prioritize their data quality over the performance of any LLM.

Generative AI is still in the early stages, and requires agencies to take an experimental approach. Although promising in its ability to create efficiencies and solve problems, generative AI may not be the appropriate tool for many use cases. Agencies should identify specific use cases and test generative AI tools in those contexts, evaluating their effectiveness on a use-by-use basis.

Data literacy training for all government employees is essential to equip them with the skills to ensure the ethical use of AI tools. Panelists discussed the ethics surrounding a scenario of employing an AI model that was knowingly trained on an insufficient data set. Ultimately, leaders will have to know how to interpret their data in order to understand if the outputs are accurate.



Learn more about Elastic’s marketing
leading solutions at
<https://www.elastic.co/>