7 Tips for Data Governance in the AI Era

As artificial intelligence (AI) gains a foothold, agencies are experimenting with how to make it work and keep it safe. What they're learning is that it all begins with data.

AI's most exciting applications come from its ability to find, absorb and analyze data. But for it to produce accurate and actionable results, the underlying data must be relevant, up to date and secure. Indeed, AI has forced agencies to step up their data governance policies, in some cases making them more sophisticated, in others enforcing what's already there.

"With the emerging technologies of AI, organizations are realizing they really do have to do the complicated, hard work of data governance and data management. It's not fun, it's not sexy. But you want to understand how reliable and accurate that data is," said Frank Garofalo, Chief Data Officer and Senior Data & Analytics Leader at Microsoft.

At a recent virtual event sponsored by Microsoft, Garofalo joined Sean Flowers, Executive Director at Ready Force Cyber, to discuss how to use data responsibly to get the most from AI. Here are their suggestions.

1 Know what you've got

The first step is to inventory the data you have across all your departments. If you have multiple databases or data specific to separate departments, you'll need to identify what can be shared. Don't forget unstructured data, such as PDFs and videos, which AI can make easier to include in analytics.

2 Ensure data integrity

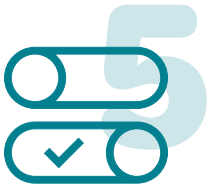
Clean up duplication and out-of-date information and establish procedures for regular updates. AI is only as good as the data that's fed into it, and answers pulled from sources that are no longer current will be inaccurate.

3 Classify and tag

Data tagging is the process of labeling data to help identify, categorize and protect it. This often involves metadata such as the author, date created, and confidentiality or proprietary status. Tagging allows AI to choose (and avoid) the right sources.

Control permissions

Some agencies have poorly controlled permissions they need to tighten up. Users may discover that AI can find data they could not or should not access, such as HR files or payroll. For instance, if the HR database is available only to HR, then non-HR AI applications shouldn't be able to reach it. A zero-trust security strategy may be useful because it gives the least privilege to everyone and then grants specific, narrow permissions where needed.



Choose which data AI can access

Agencies should have constraints on where many AI applications can access data. For instance, you may not want an AI search to go out to the wider internet, or you may want to confine a chatbot to answers from a specific public database. "Grounding" AI will restrict it to internal data and improve the accuracy and relevance of its results.

Head off "shadow IT"

AI is an attractive tool, and employees want to use it. But if your access policy is too strict — for example, blocking any use of AI — some workers will circumvent your controls and go to insecure public versions, which leads to "shadow IT." Agencies should establish data and AI policies that allow users to do their work without resorting to workarounds.

Educate your users

Have an AI policy that employees understand and agree to. Emphasize that anything entered in public-facing AI is available to the world, much like social media. Let them know that putting agency data into a public AI model means the model will retain and use it as part of its training, running the risk that it will reappear in AI responses to other, non-agency users.

"It's security vs. privacy vs. functionality. So, what's more important to you? That depends on the person. But for organizations, you have to be able to classify your data, control access and then make sure people understand your use policy."

— Sean Flowers, Ready Force Cyber

govloop

Watch the full event on demand

Microsoft