Addressing the siren call of generative AI

How government agencies can protect themselves from generative AI risk

Many emerging technologies have been labeled as "disruptive"—artificial intelligence (AI), machine learning (ML), 5G, internet of things (IoT), and edge computing, for example—but none has more potential to transform how we *all* live and work than **generative AI**. Generative AI models can produce content such as articles, essays, poetry, software source code, images, and even videos on practically any subject based on simple human prompts. They can achieve remarkably human-like results with little or no human intervention. And there's almost no barrier to entry; all you need is an internet connection.

It's difficult to escape stories of these highly capable generative AI models, including ChatGPT and DALL-E—they've been grabbing headlines for months. The record-setting pace of ChatGPT's adoption alone is evidence of its potential impact: it reached more than 100 million users within two months of its launch[1], compared to the four and a half years it took Facebook or the two-plus years it took Instagram[2]. ChatGPT's parent company, OpenAI, reports it now has over 1 billion visits per month[3]—a number that's growing so rapidly it's likely well out of date by the time you read this.

Generative AI isn't some fad or far off disruption. Millions of private sector organizations and individuals have already started using generative AI for content creation for websites, social media posts, research papers, cover letters, emails, text summarization, and more. It's real, it's here, and it's something nearly every agency at all levels of government must address—now.

Why modern government is important

Government agencies in the U.S. must modernize in order to keep up with changing user needs, regulations, and health and public safety requirements. Leaders of modern governments rethink business processes and service delivery models to more effectively achieve their mission. This article is one of a series that features how modernizing affects the government workforce and the user experience, improves security and public trust, and accelerates the digital journey. KPMG team members offer insights intended to help guide governments in their modernization efforts to encompass all processes, technologies, policies, and the workforce so each works together to create connected, powered, and trusted organizations.

[1] Source: Reuters, "ChatGPT sets record for fastest-growing user base—analyst note," Krystal Hu, February 2, 2023
[2] Source: The Motley Fool, "The Social Media Platforms That Hit 100 Million Users Fastest," Natalie Walters, April 20, 2019
[3] Source: Similarweb website, April 2023

Consider the risks

Because of the democratization of generative AI, all public sector organizations, no matter their size or technological maturity, need to start thinking about a host of risks surrounding it, both internal risks (those related to the use of generative AI within your organization) and external risks (those related to others using it to interact with you).

These risks can run the gamut, including privacy, trust/ethical, financial, reputational, compliance, security, algorithmic discrimination, and fraud. A few of my colleagues have written a great article on many of these risks, at least as they apply to private sector organizations, so I won't repeat them here. But let's consider a couple examples that are more specific to government agencies.

Given its amazing capabilities, generative AI has many potential applications within government. But before you begin envisioning how you might exploit it, it may be more important to recognize how others might be using it to exploit you.

With generative AI, inundating a system with seemingly genuine content has never been easier. Think about the effect this could have on public comments. Anyone who's sufficiently motivated can use a generative AI model to ghostwrite a virtually unlimited number of comments, thereby flooding the system with their opinion while drowning out actual human voices. If you're on the receiving end of this deluge, how do you know which voices to listen to? Which submissions were written by humans? Which by AI?

The same is true if you're relying on external data as input for a model used to support decision-making. If a malicious actor decides to put their metaphorical finger on the scale or poison the data well, it can be difficult to identify such interference—and therefore to know if your decisions have been influenced. Less sophisticated types of fraud can also benefit from generative AI. Need to craft a clever phishing campaign? Use generative AI to write email hooks. If even one employee falls prey to such a scam, your agency may be at serious risk.

A very long tail

The risk isn't necessarily malicious, though. It could be inadvertent. For example, many have observed that the tone of ChatGPT's output is very confident, and that it can generate very professional-looking and accurate-sounding content that lacks any factual basis, including entire data sets, and realistic charts and graphs. Someone submitting a government grant application or a response to request for proposal (RFP) for a government contract, or providing compliance or progress updates to one that had already been awarded, may be using a generative AI model to complete some of the more perfunctory sections.

Malicious or not, what if generated content with no factual basis makes its way into more significant sections? What if it makes a claim in an RFP, for example, about the properties of a military helmet that sounds accurate but isn't? More importantly, what if no one catches it? Consider how long the tail of such fabricated content could be. Data submitted in this way becomes the property of the government—valuable insights that are often leveraged multiple times for many other purposes, such as training government AI models or providing input to them, setting future RFP requirements, or making future grant or contract award decisions.

When leveraged for these purposes, such data may not be treated with the same caution or skepticism usually reserved for "external" data sources—databases that are outside of government, government control or governance. Despite its origin, this is likely to be considered government data, and may be relied upon to have been governed and vetted as such. Many organizations outside of government may use it, too, with the same level of trust. The damage could be widespread and lasting.

Addressing risks with DataOps

What's the solution to these generative AI challenges? More AI! Even if you don't use generative AI, you may need other AI to detect and defend against it.

At present, the defense against external risks might appear to be lagging behind the offense. For instance, to detect content generated by ChatGPT, OpenAI has released AI Text Classifier. Yet it has just a 26 percent accuracy rate[4]—worse than flipping a coin. Given the pace of everything generative AI, however, this is likely to change soon.

Even without a new tool designed specifically to detect generative AI output, you still have other tools in your bag. For years, data teams have been using machine learning and other methodologies to detect anomalies in their data, including data from external sources, under the auspices of a broader discipline known as **DataOps**.

DataOps is a set of practices, processes, and technologies that applies a product mindset to the design and delivery of data products. Inherently, data quality and integrity are paramount features of these data products. As the name implies, DataOps extends the DevSecOps approach from software development to data, enhancing the quality, speed, security, access, and continuous improvement of data products. Effective DataOps can help automatically identify data anomalies throughout the data lifecycle to help prevent such outliers from negatively affecting model output—including anomalies that may originate from generative AI.

Addressing risks with responsible AI

Mitigating AI risk doesn't end with data. The entire ecosystem in which the models are developed, deployed, and used must also have a set of rigorous practices, processes, and technologies designed to address AI risk. That discipline is **responsible AI**.

Responsible AI is an approach to designing, building, and deploying AI systems in a safe, trustworthy, and ethical manner so that government agencies can accelerate value for constituents, organizations, and society with confidence. KPMG has developed eight core principles that guide our approach to responsible AI across the AI/ML lifecycle:

- **Fairness** — AI models should be equitable and free from bias
- **Explainability** —Models should be understood, documented, transparent, and open for review
- **Accountability** — There should be mechanisms to drive responsibility across the model lifecycle
- **Data integrity** — There should be steps focused on data quality, governance, and enrichment to engender trust
- **Reliability** — AI systems should perform at a desired level of precision and consistency
- **Security** — Safeguards should be in place to defend against unauthorized access, corruption, or attacks
- **Privacy** — Compliance requirements for privacy, regulations, and consumer data usage should be adhered to
- **Safety** — AI should not negatively impact humans, property, or the environment.

[4] Source: OpenAI website, "New AI classifier for indicating AI-written text," Jan Hendrik Kirchner et al., January 31, 2023

In plain English, responsible AI means engineering privacy, transparency, accountability, and explainability into the entire AI ecosystem. For example, it's not enough for an AI system to automatically label an incoming public comment as AI-generated. That system also needs to provide a human-interpretable explanation of that decision, providing stakeholders with the opportunity to review that decision and agree or disagree with it. Should they disagree, the system needs to remember their decision so that it is continuously learning and improving. Engineering such a feedback loop is part and parcel of DataOps today.

It's clear that the AI ecosystem should be carefully reviewed and continuously monitored to help achieve the eight principles of responsible AI, but the question is who will do it? A recent KPMG survey of 140 U.S.-based executives from organizations with revenue of $1 billion or more[5] revealed that 84 percent believe that independent audit of their AI models will be a regulatory requirement within the next one to four years. However, only 19 percent said that they explicitly have the expertise to conduct such audits internally. If private sector organizations of this size are struggling, what does that mean for those in the public sector?

As you might expect, KPMG has invested in both its DataOps and its responsible AI capabilities. With our extensive experience in AI, data and analytics, engineering, and software development, we have the necessary breadth and depth of capabilities designed to help you use AI responsibly and effectively, at scale, to achieve tangible mission outcomes.

believe that independent audit of their AI models will be a regulatory requirement within the next one to four years

said that they explicitly have the expertise to conduct such audits internally

In conclusion

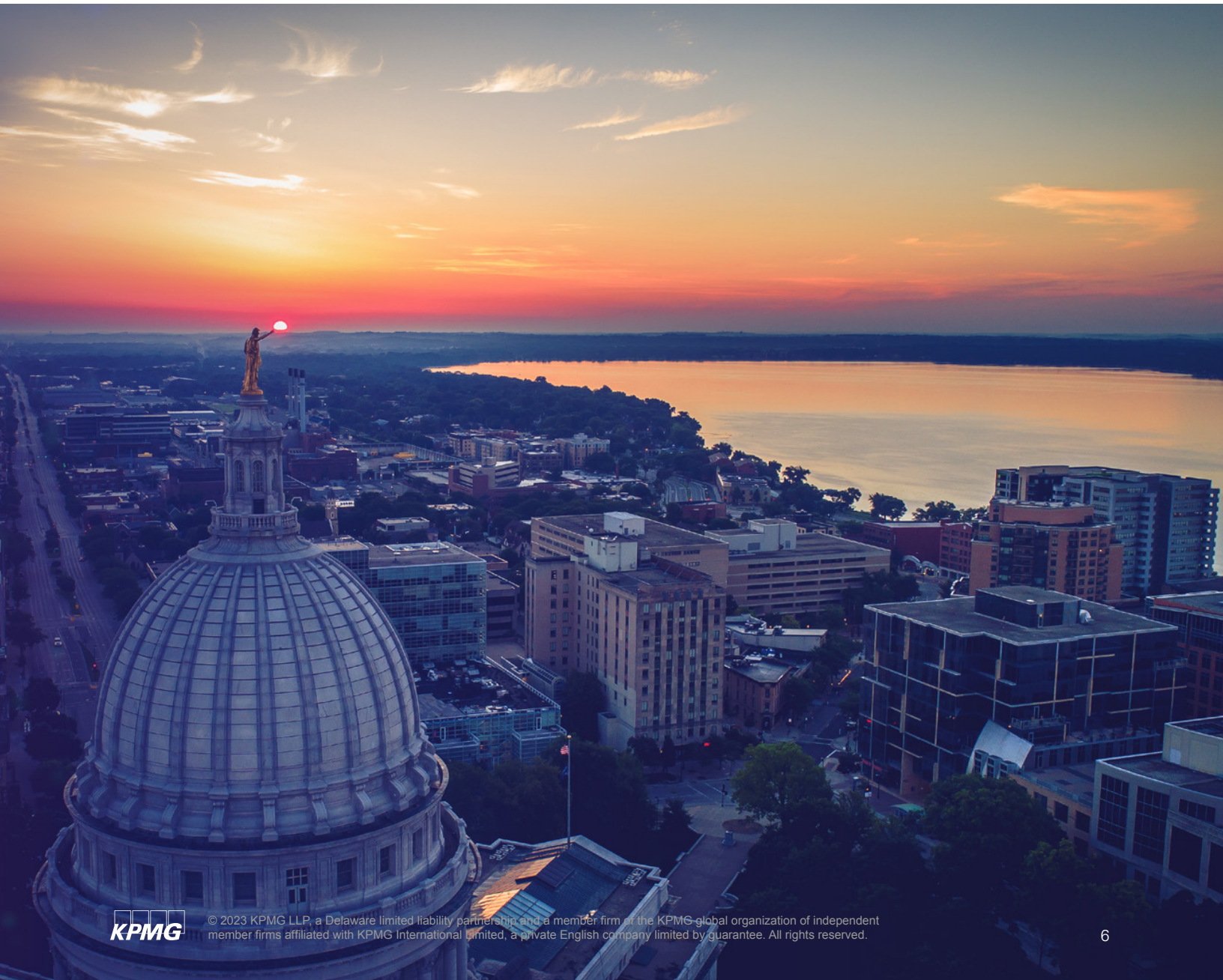
The cat is out of the bag with generative AI. Given its allure, people are going to use it, whether you try to prevent them or not—as we see at many large organizations. In this brave new world, organizations need to educate the workforce and incorporate a responsible AI approach to design, build, and deploy AI systems in a safe, trustworthy, and ethical manner so that they can accelerate value for their stakeholders with confidence.

[5] Source: KPMG Artificial Intelligence Risk survey, September 2022

About KPMG

KPMG has worked with federal, state, and local governments for more than a century, so we know how agencies work. Our team understands the unique issues, pressures, and challenges you encounter in the journey to modernize. We draw on our government operations knowledge to offer methodologies tailored to help you overcome these challenges and work with you from beginning to end to deliver the results that matter.

The KPMG team starts with the business issue before we determine the solution because we understand the ultimate mission. When the way people work changes, our team brings the leading training practices to make sure your employees have the right knowledge and skills. We also help your people get value out of technology while also assisting with cloud, advanced analytics, intelligent automation, and cybersecurity. Our passion is to create value, inspire trust, and help government clients deliver better experiences to workers, citizens, and communities.



Contact

Viral Chawda
Principal, Advisory
Head of Technology Practice—Government
KPMG LLP
832-535-8712
vchawda@kpmg.com

---

read.kpmg.us/modgov