IDC

SPOTLIGHT

Sponsored by: Anonos

Leveraging personal data to drive insights for enterprise activities is fraught with peril in the current data privacy landscape. GenAI adds another layer of complexity. Organizations will need to look to performant privacy to safely seed GenAI models.

Variant Twins: The Key to Safely Leveraging Data for AI

November 2023

Written by: Ryan O'Leary, Esq., Research Director, Privacy and Legal Technology

Introduction

Digital transformation has dramatically increased the amount of data being generated and used in the enterprise today. In IDC's 2022 *Data Privacy Survey*, 68% of organizations reported anticipating data volume increasing over the next three years. Of that number, 40% expect the data volumes to increase by 25%+.

Early this year, even before the most recent upsurge in interest, 61% of organizations were at least experimenting with generative AI (GenAI), according to IDC's *Future Enterprise Resiliency and Spending Survey, Wave 2.* GenAI and other types of AI require massive volumes of data to train their algorithms. While these technologies are table stakes for organizations that want to be competitive, there is a significant risk of disclosure of sensitive corporate information.

AT A GLANCE

KEY TAKEAWAYS

- Risk associated with leveraging sensitive data is untenable. GenAI's impact will be curtailed if there is no data available to train the models.
- Performant privacy technologies and variant twins alleviate the risk and enable utility at scale of organizational data.
- Security and privacy are the two biggest roadblocks to GenAI for enterprises.

The tension between the benefits and the risks of using GenAI, as well as other types of AI, requires a new technological approach to mitigate the security and privacy risks currently impeding greater enterprise use of these technologies.

Data is used to underpin business operations and customer applications. It's collected and used to fine-tune the customer experience and make things easier. There is even data about data — consider the number of security organizations that are actively collecting security event data to monitor and reduce risk. The risk is untenable and organizations will not be able to seed GenAI models without appropriate safeguards. To leverage GenAI, organizations will need to shift their privacy left and embrace new controls and frameworks — frameworks that leverage performant privacy and variant twins. These variant twins support organizations in aligning with business and compliance requirements throughout their data journey by enabling a "Shift Left Privacy" strategy of embedding technical controls into data early in the process.

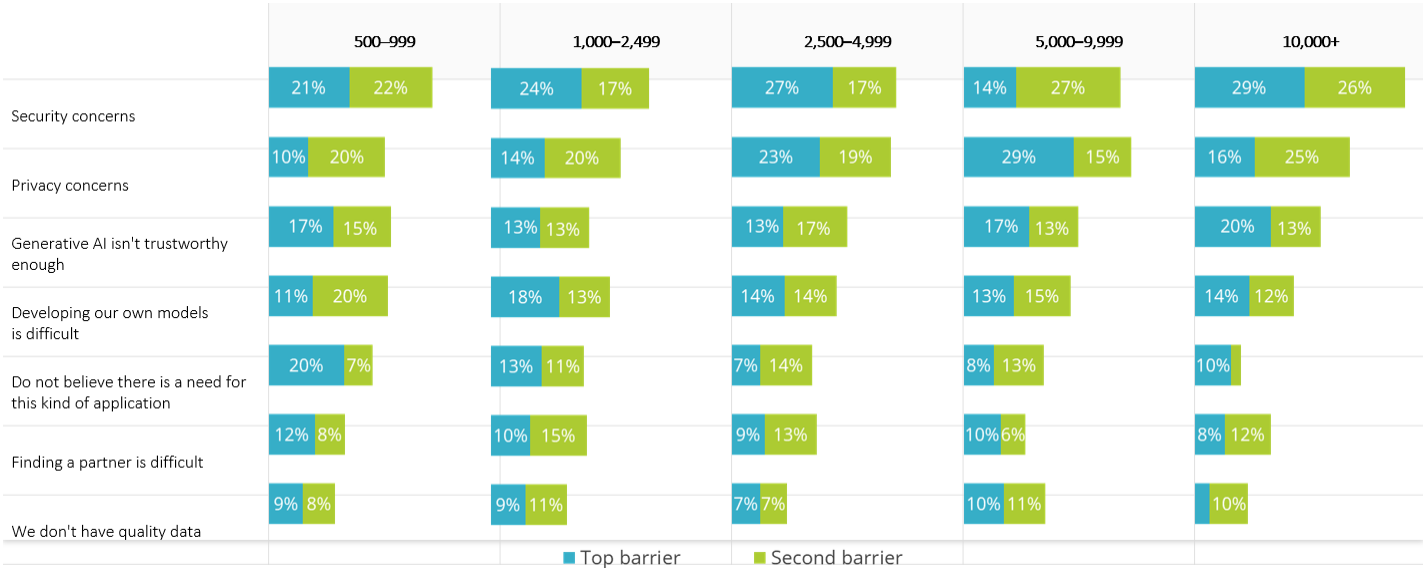
The need for a new framework underscores the reality that if adequate security and privacy solutions were already in place, the current hesitancy toward employing sensitive business data in GenAI and AI would not be a concern. Organizations will need to find solutions to be able to leverage the data while still protecting corporate trade secrets and IP and preserving the privacy of individuals. Enter performant privacy and variant twins.

Performant Privacy Safeguards Enable Deriving Value from Data

Generative and other types of AI cannot provide results without data to seed their algorithms. AI and data processing will be table stakes for organizations that want to be competitive. These organizations will need to consider embedding performant privacy into the very systems that contain their data. This is one of the few ways that organizations can protect the data while also being able to use it. When it comes to GenAI, security and privacy are the two biggest concerns for enterprises (see Figure 1).

FIGURE 1: ***Privacy and Security Concerns Are Top of Mind for Organizations Exploring Generative AI***

Q What are the top 2 barriers to using generative AI in your organization (top versus second)?



n = 890 (500–999: 137; 1,000–2,499: 328; 2,500–4,999: 175; 5,000–9,999: 135; 10,000+: 115)

Source: IDC's Future Enterprise Resiliency and Spending Survey, Wave 6, July 2023

Performant privacy-based controls reconcile tensions between ensuring high data utility and effectively protecting data when in use between and outside of perimeters. IDC defines "performant privacy" to describe the state of the art for retaining the performance characteristics of unprotected source data (such as speed of processing, accuracy, and fidelity), while delivering privacy that satisfies compliance requirements. While inherent privacy issues have existed with AI for some time, generative AI and large language models (LLMs) are amplifying the risk and opportunities of ML because they dramatically democratize access to the insights derived from ML. Performant privacy can protect against prompt leakage of proprietary context or data and can safely train domain-specific LLMs.

Using variant twins, a data controller can create privacy-secured data sets from which it is impossible to re-identify individuals. Variant twins essentially remove risk from data and allow it to be used safely.

Performant privacy-based controls are one of the few tools that reconcile data protection and utility. Performant privacy unlocks data flows, and reduces data compliance risks, thereby increasing

data value without impeding timely access to and use of data for accelerated decision-making and business value. Performant privacy baked into the tools where the data exists solves probably the gnarliest problems of data privacy.

Considering Anonos

Anonos Data Embassy software uses performant privacy-based controls to reconcile tensions between ensuring high data utility and effectively protecting data when in use between and outside of perimeters. Benefits of the solution include the ability to leverage preapproved data sets that significantly limit risk to the organization. Anonos has the ability to create protected versions of data to enable protected use. Anonos Data Embassy software leverages multiple data protection techniques to create new, compliant, and approved use case–specific protected versions of data called "variant twins." Anonos does not rely on just one method of protection.

A variant twin essentially creates a duplicate but differentiated data set that mimics the characteristics essential to driving value from the data without exposing the underlying data. The underlying data can only be extracted with an encryption key or via additional information that is held separately and securely. With Anonos variant twins, a data controller can create privacy-secured data sets for generative AI from which it is impossible to discern corporate trade secrets or IP or re-identify individuals. Variant twins essentially remove risk from data and allow it to be used safely. Organizations can have their cake and eat it too.

The variant twins remove the obstacles faced when trying to process data or seed AI models. The data is protected during computation to allow for more accurate results than traditional synthetic or dummy data would provide. Some of the benefits include:

- **Safeguarding sensitive data.** As enterprises explore various use cases for GenAI, they may need to work with sensitive or proprietary data. Performant privacy ensures that this data remains protected during active use, allowing for more comprehensive experimentation without compromising data security.
- **Facilitating collaboration.** Many GenAI projects involve collaboration between multiple departments, partners, or even different organizations. With performant privacy in place, data can be shared and used more freely without the fear of exposing sensitive information, enabling more extensive and diverse collaborations.
- **Maintaining data quality and integrity.** Performant privacy ensures that while data is protected, its quality and integrity are not compromised. This is crucial for GenAI models, which rely on high-quality data for accurate outcomes. Enterprises can thus be confident in the results of their explorations.
- **Regulatory compliance.** With increasing data privacy and protection regulations, enterprises often face challenges in using data for AI projects. Performant privacy can ensure that data usage aligns with regulatory standards, allowing for smoother exploration of GenAI use cases without legal hindrances.
- **Enhancing stakeholder trust**. Customers, partners, and other stakeholders are more likely to support and engage in GenAI projects if they are assured of data privacy. Performant privacy bolsters this trust, facilitating more open discussions and exploration of potential use cases.

Anonos also has a synthetic data generation capability. Organizations can artificially generate data that mimics production data's structure and statistical properties. Synthetic data can be used for testing and training, eliminating privacy risks in software and analytical model development.

Anonos is enabling data modernization by helping organizations turn privacy and compliance roadblocks into on-ramps to digital business.

Challenges

It will take time before organizations start to truly deploy GenAI. Organizations are acting with caution, and Anonos will need to break through the knowledge gap. There is a knowledge gap in the market about Anonos and what the company does. Anonos has a big task to educate the market on how its implementation of performant privacy-based controls is different and what value it adds to ensuring compliant sharing and consumption of data at all levels of sensitivity.

Conclusion

As the risk of leveraging data within GenAI and LLMs continues to explode, organizations will need to find ways to safely collect and process data. The resources and roadblocks associated with managing dynamic consent frustrate the ability to leverage and get value out of data. Performant privacy technologies like Anonos variant twins will need to be embedded across the business to obfuscate the underlying data and allow the utility of processing without the risk.

About the Analyst

Ryan O'Leary, Esq., *Research Director, Privacy and Legal Technology*

Ryan O'Leary is a research director in IDC's Security and Trust research program covering privacy and legal technology. In this role, Mr. O'Leary leverages his legal experience to provide perspective on changes in laws, shifting regulation, and other market forces that affect technology decision-making today for both law firms and corporations. He also provides thought leadership that technology suppliers and technology buyers may use to develop effective strategy for the future. Mr. O'Leary's core research coverage includes the evolution of ediscovery and legal technology, as well as the evolution of privacy compliance technology and impacts of new and emerging data privacy regulation.

MESSAGE FROM THE SPONSOR

Anonos delivers the competitive benefits of sharing, combining, and using sensitive data for AI, which data is otherwise off-limits without protecting it when in use wherever it travels.

- Anonos does this by enforcing performant privacy controls that ensure that the inherent characteristics of original, unprotected data, such as processing speed, accuracy, and fidelity, remain intact, all while ensuring compliance with the most rigorous global privacy and security standards when the data is in use.
- Our Variant Twins secure data during processing wherever it flows — within an organization, its broader network, and across organizational and jurisdictional borders.
- By merging data protection with utility, our globally patented Data Embassy software amplifies the value of high-risk data (like PII, personal data, and trade secrets) to pave the way for innovative new AI-driven insights and revenue.

Learn more about empowering faster and compliant use of AI to improve business outcomes at www.anonos.com.

IDC Custom Solutions

The content in this paper was adapted from existing IDC research published on www.idc.com.

IDC Research, Inc.
140 Kendrick Street
Building B
Needham, MA 02494, USA
T 508.872.8200
F 508.935.4015
Twitter @IDC
idc-insights-community.com
www.idc.com