Enabling Enterprise Use of AI Without the Risk

Q: What are the biggest barriers to enterprise use of AI?

A: Lack of adequate security and privacy are the two biggest concerns for enterprise use of AI. Generative and other types of AI cannot provide results without data to train the models. While AI is essential for organizations to be competitive, enterprises are concerned about disclosing sensitive data (e.g., personally identifying data, corporate trade secrets, and intellectual property) when using these models. A new approach to reducing the sensitivity of data and removing the obstacles to compliant utility is necessary so that more data is available more quickly to complete AI projects lawfully. In this manner, data-driven initiatives can be expanded and expedited in compliance with applicable regulations and corporate goals and objectives.

Q: How do you overcome these obstacles while preserving utility?

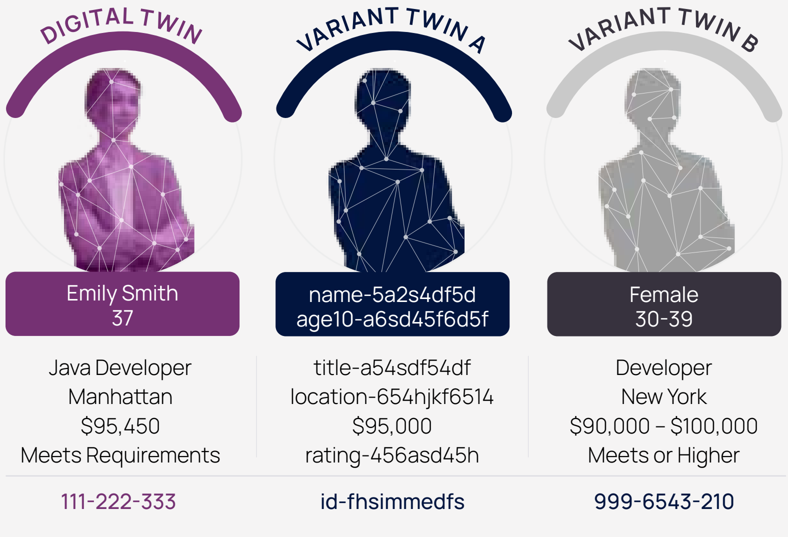
With Variant Twins®. Variant Twins are protected outputs engineered with collaboration between the privacy/legal and data teams to meet the needs of specific use cases based on preconfigured, multi-level privacy and security controls that are technologically enforced. Any data, regardless of how sensitive it may be – everything from personally identifying data to trade secrets or IP – can be transformed into a Variant Twin that is safe for use anywhere without loss of accuracy or speed. The full-spectrum protection Variant Twins provide means faster time to enterprise insight because legal and business stakeholders can collaborate to increase the use of valuable, sensitive data.

Q: What is a Variant Twin?

A: If you work with data, you've likely heard the term Digital Twin to refer to a digital representation of a person, place, or thing. However, this representative model contains sensitive data that creates privacy and security concerns, limiting use and value. In contrast, a Variant Twin is source data that has been transformed into a new protected data asset. **Variant Twins reveal only the minimum identifying information necessary to deliver the original data's underlying accuracy and utility to meet use-case needs without the risk, in compliance with data privacy laws.** A Variant Twin is also portable, meaning it can be processed safely anywhere – on a local desktop or in the cloud – for use with any application. And, Variant Twins do not result in liability or disclosure obligations if breached.

Each Variant Twin contains a subset of the source digital twin data, comprising only the data elements that are needed and authorized for a particular use or analysis.

Because use-case-specific Variant Twins originate from source data, downstream data fidelity is maintained, even when privacy protections have been applied.



DIGITAL TWIN	VARIANT TWIN A	VARIANT TWIN B
Emily Smith 37	name-5a2s4df5d age10-a6sd45f6d5f	Female 30-39
Java Developer	title-a54sdf54df	Developer
Manhattan	location-654hjkf6514	New York
$95,450	$95,000	$90,000 – $100,000
Meets Requirements	rating-456asd45h	Meets or Higher
111-222-333	id-fhsimmedfs	999-6543-210

> Anonos' technology allows for re-linkable non-identifying personlized data showing how advances in technology make our ability to leverage data for better outcomes plausible today.
>
> **Nadia Hewett**
> Lead for the Data for Common Purpose Initiative (DCPI), World Economic Forum

**Sound too good to be true?
Take the 'Variant Twin Challenge' to see for yourself.
LearnMore@anonos.com**

Q: How are Variant Twins created?

A: Variant Twins are created with the Anonos Data Embassy® platform, using a combination of state-of-the-art privacy-enhancing and de-identification techniques. Connected to the desired source, the software transforms the selected data into a new asset based on pre-configured rules and automated workflows that determine what level of identifying information is revealed, to whom and for what purpose. The protected, use-case-specific Variant Twin is then made available for processing and analysis without violating privacy, security or other compliance requirements. Built-in protection is effective against single- and multi-use identity disclosure risk because direct and indirect identifiers are delinked from each other and the source data.

Q: What are the unique advantages of Variant Twins for AI and analytics?

A: Anonos spent 10 years researching and developing Variant Twins to meet the statutory requirements of the General Data Protection Regulation (GDPR) and other data privacy laws. However, they're future-proof because they're generated from a toolbox of protections according to local rules that are continuously enforced wherever they travel. One Variant Twin can embody a combination of numerous data privacy and security techniques.

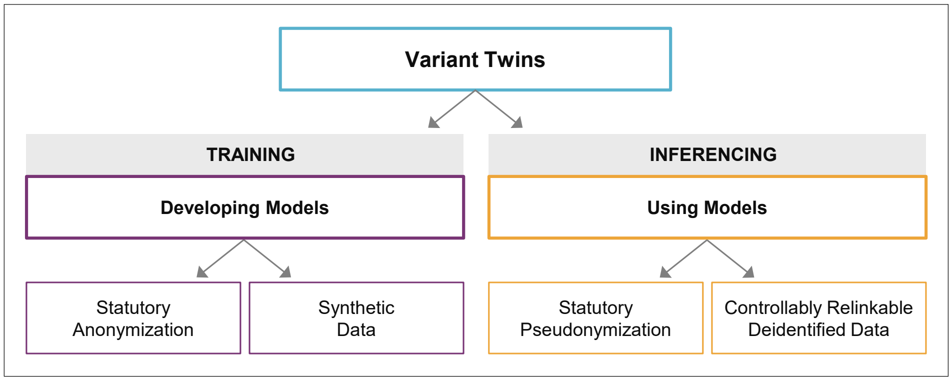
Guardrails Are Not Enough

While encryption, access controls, masking, and tokenization act as "guardrails" to curtail specific privacy vulnerabilities, they fall short of precluding unauthorized disclosure of sensitive data when using AI. The diversity of datasets used in LLMs and multiparty AI projects opens many avenues for drawing correlations between seemingly innocuous information to reveal sensitive information (i.e., the "Mosaic Effect"), even in the presence of these guardrails.

To illustrate, access controls and encryption serve to safeguard data against unauthorized access, but they fail to restrict entities with authorization from exploiting the data, including using the data to reveal identity via the Mosaic Effect. Masking and tokenization protect data in isolation; nonetheless, the emergence of the Mosaic Effect results from integrating and correlating diverse, ostensibly "anonymized" datasets, unveiling identities and patterns that are imperceptible within any single data set.

Developing Models: Statutory Anonymization with Synthetic Data

Training models in an AI world while protecting sensitive data requires anonymous data. Synthetic Data overcomes the limitations of guardrails as noted above because it represents a fundamentally different approach to anonymizing data. With Synthetic Data, artificial intelligence (AI) and machine learning (ML) algorithms are used to capture the statistical relationships in a data set as a mathematical model. That new model is then used to generate entirely new records that preserve the analytic utility in the data, but no newly created synthetic record is mappable to an original record. Therefore, synthetic data can be thought of as being born anonymous instead of transforming cleartext data to *become* anonymous. **Variant Twins enable Synthetic Data for training AI models.**



Using Models: Statutory Pseudonymization with Controllably Relinkable DeIdentified Data

Since Synthetic Data makes de-identification irreversible, Variant Twins enable the use of Statutory Pseudonymization to preserve accuracy and enable authorized relinking only under controlled conditions. This technique was first defined under the GDPR and is being adopted in many jurisdictions. It protects data by transforming it to prevent reidentification by anyone other than the data controller without reducing accuracy or utility. Its five distinguishing features are 1. Protection for all data elements; 2. Protection against singling-out attacks; 3. Dynamism (i.e., using different pseudonyms at different times for different purposes); 4. Selective use of non-algorithmic lookup tables; and 5. Controlled re-linkability. The level of data protection Statutory Pseudonymization provides is superior to anonymization because controlled re-linkability allows for the application of protection to many more fields without destroying the utility of the protected data set. **Variant Twins enable Controllably Relinkable Deidentified Data for using AI models.**

Q: Do Variant Twins leverage synthetic data?

A: Variant Twins use synthetic data to create test data and train artificial intelligence and machine learning models in a way that overcomes the limitations of using synthetic data by itself. With Data Embassy, you can protect data across the full spectrum of use cases – from testing through production – dialing up or dialing down the level of privacy preservation required based on the use case.

**Read IDC Spotlight Report-
Variant Twins: The Key to
Safely Leveraging Data for AI
at: anonos.com/AI-Spotlight**