SHARDSECURE

Solution Brief

Protection for AI/ML Models & Training Data

Learn how ShardSecure® protects high-value AI/ML models and training data against common threats like industrial espionage, ransomware, tampering, and outages.

Hundreds of millions of dollars: the high cost of developing AI/ML models & training data

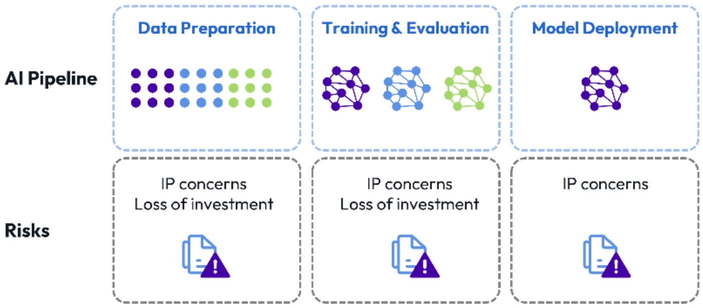
As the artificial intelligence sector grows, the true value of AI/ML models and training data is becoming clear. These datasets serve as the foundation for new technologies, allowing companies to create accurate and effective large language models, generative AI models, and machine learning algorithms. With high-quality training data, businesses can build robust AI/ML models that accomplish everything from better customer targeting and fraud detection to reliable medical diagnoses and compelling artwork.

AI/ML models and training data are proprietary and unique, containing valuable insights and information. Companies are dedicating substantial time and resources — in some cases, hundreds of millions of dollars — to collect, clean, and organize large volumes of data. Processes like data acquisition, data cleaning, data labeling, and data augmentation all contribute to the overall cost of these models, as do infrastructure investments like high-performance compute and cloud storage. As the intellectual property of the future, these assets represents a real competitive advantage in the market.

Given their R&D cost and susceptibility to tampering and scraping, AI/ML models and training data must be protected as a trade secret and safeguarded from unauthorized access. ShardSecure offers an innovative software solution to protect AI/ML models and training data, with advanced data security and privacy, support for data sovereignty, robust data resilience, and agentless integration.

Protection for AI/ML models & training data with ShardSecure

AI/ML models and training datasets are increasingly on the radar of industrial espionage groups and ransomware attackers. Their importance to companies and the ease with which they can be compromised make them a high-value target.



	Data Preparation	Training & Evaluation	Model Deployment
AI Pipeline			
Risks	IP concerns Loss of investment	IP concerns Loss of investment	IP concerns

AI/ML models and training datasets are also maintained in the cloud with services like AWS S3, Azure Blob Storage, and GCP Storage. While this storage can facilitate lower costs, it also brings data security risks. The threat of data exfiltration, the likelihood of human error, and the undeniable fact of cloud admin access to sensitive data all mean that companies must take additional security measures to keep their valuable models and training data safe.

ShardSecure's Microshard platform offers an innovative solution to protect AI/ML models and training data in the cloud, including in multi-cloud architectures. The ShardSecure platform provides advanced data security, allowing organizations to safeguard their assets from malicious actors without the complexities and performance drawbacks of traditional solutions.

Separation of duties for data sovereignty

The ShardSecure platform allows companies to control exactly who has access to their AI/ML models and training data. As an innovative approach to file-level encryption, the platform separates infrastructure administrators and cloud service providers from access to sensitive data.

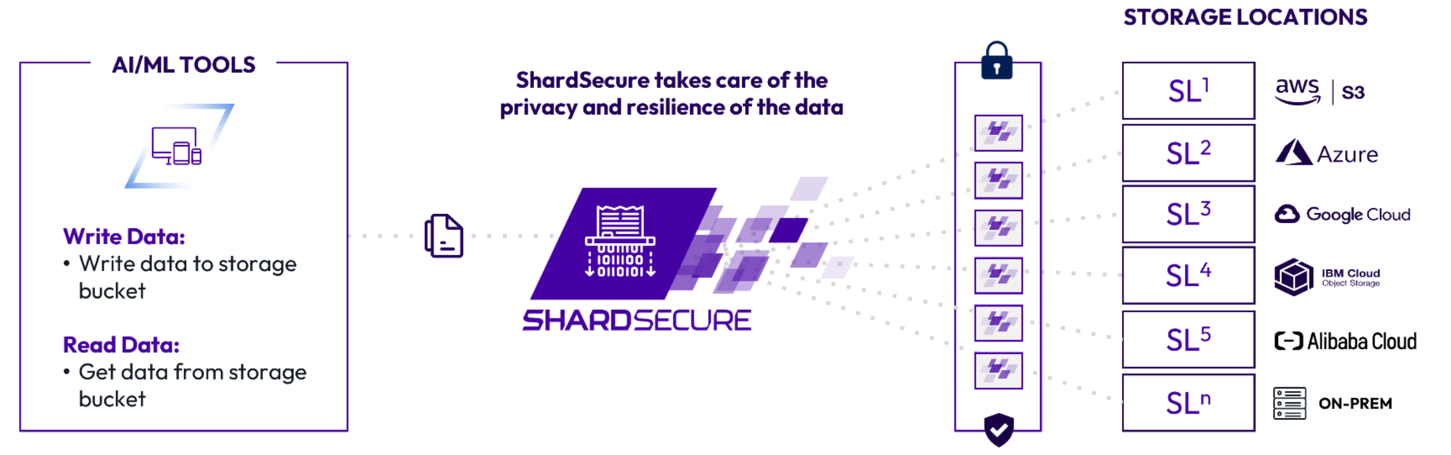
ShardSecure's approach also ensures data sovereignty, as it allows companies to solve residency issues and control exactly where they store their data in the cloud.

ShardSecure mitigates the impact of data tampering by detecting any unauthorized modifications. If an organization's training data is found to be inaccessible, deleted, or tampered with in an adversarial machine learning attack, the platform can reconstruct the affected datasets without the need to restore them from backups. This self-healing feature works automatically and transparently to preserve data integrity.

ShardSecure also ensures high availability for AI/ML models and training data. Each instance of ShardSecure is a virtual cluster, and customers can configure 2+ virtual clusters for failover.

Simple, agentless integration and strong performance

The ShardSecure platform provides an easy and transparent implementation with no need to change AI/ML tools or data flows. Without the overhead and complexity of traditional data security solutions, it is vendor-agnostic and appears to applications as a storage location. It also allows for instant data access and fast data migration with just a few clicks.



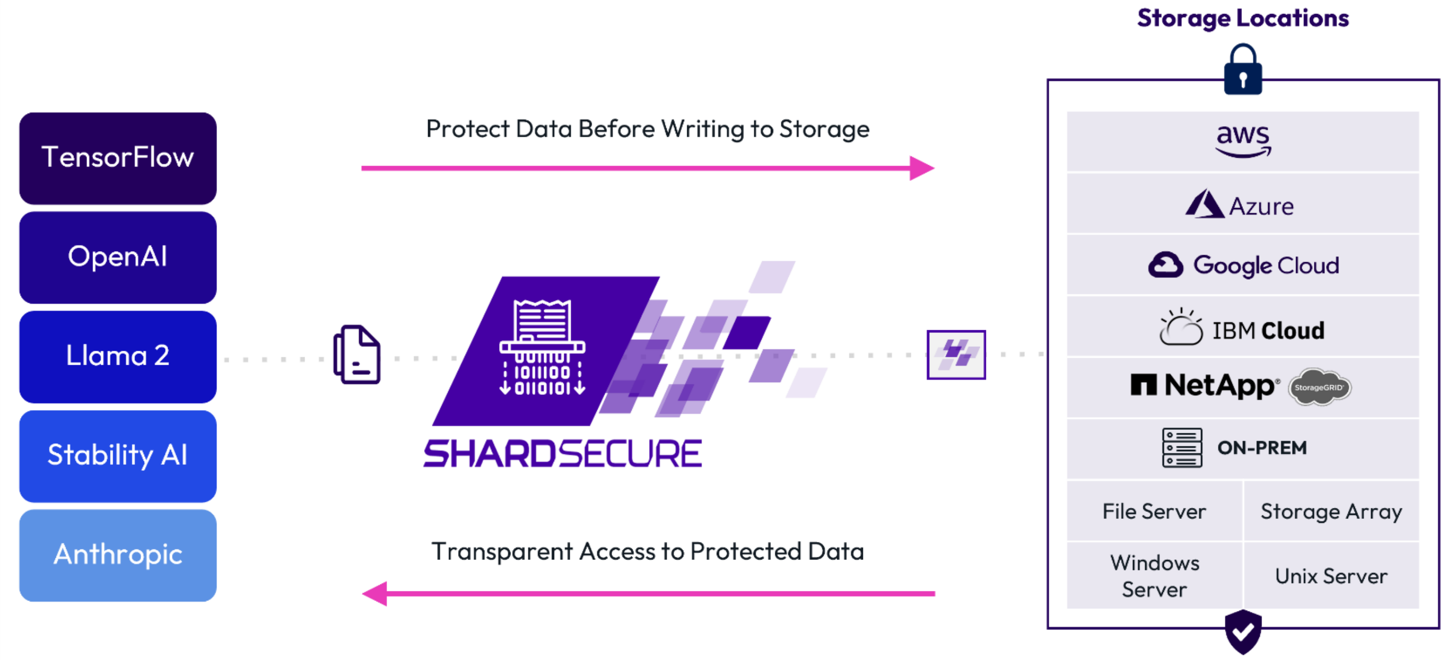
By protecting data before it's stored, performing regular integrity checks, and offering features like object locking and immutable storage, the ShardSecure platform ensures that AI/ML data cannot be scraped by unauthorized users.

Robust data integrity and high availability

AI/ML models and training datasets are highly sensitive to data tampering. In what's known as adversarial machine learning or data poisoning, a malicious hacker can change the outcome of a model simply by adjusting sets of training data to mis-teach or bias the model. This kind of data manipulation is on the rise and can render a previously trained model ineffective.

The S3-compatible API and iSCSI interface make it simple for most AI/ML tools to migrate to ShardSecure with minimal configuration changes. ShardSecure's native multi-cloud and hybrid-cloud support provides a single interface to manage storage locations and move data as needed without impacting application performance. The platform also uses scale-out architecture and parallel I/O to achieve performance on par with native object stores.

As a result, ShardSecure has a minimal impact and works in the background as a transparent, zero-downtime event. Data protection is achieved without expending significant resources on running and maintaining complex systems.



Learn more

ShardSecure integrates seamlessly with existing security controls and cloud storage providers for ease of deployment. In addition to protecting AI/ML models and training data, we support robust data resilience, cloud ransomware mitigation, and regulatory compliance.

To learn more about our technology, follow us on social media or visit us online.

@ShardSecure
@ShardSecure
@ShardSecure

101 Avenue of the Americas
9th Floor, New York, NY 10013
United States of America

info@shardsecure.com

SHARD SECURE