Solution Brief

THALES

CYBERSECURITY

CipherTrust Tokenization Solutions: Secure AI Without Slowing Innovation

cpl.thalesgroup.com

THALES

CYBERSECURITY

When you need

- **AI deployment without exposing regulated data during prompts or retrieval**
- **Protection against sensitive data persisting in logs, embeddings or AI-generated outputs**
- **Continuous PCI DSS 4.0 and privacy compliance across AI-driven workflows**
- **Secure data before it enters generative AI (GenAI), foundation models and AI agents without expanding data-in-use risk**

The Challenge

Executives today are under immense pressure to deploy AI to maintain a competitive edge. However, AI Chatbots and Large Language Models (LLMs) create a unique data security crisis: once sensitive data enters a prompt, it can be absorbed into the model's memory or logs. Achieving AI-driven innovation without creating new vulnerability gaps is the modern executive's primary roadblock.

A single AI interaction can trigger retrieval across multiple systems, generate derivative artifacts, create new data records that can be written to a database, and produce outputs that are reused elsewhere in the organization. AI systems do not simply access data; they replicate, transform, summarize, embed, and reuse it.

Once sensitive data propagates as cleartext, governance becomes exponentially more complex, and in some cases, irreversible.

From both security and privacy perspectives, the question is no longer, "Who can access the data?" The question is "Where does the data propagate and can we still govern it there?"

How Thales Can Help

Thales CipherTrust Tokenization solutions enable organizations to "blindfold" AI systems, so that business analytics run unchanged while sensitive data is removed before crossing either of the two critical AI exposure boundaries:

1. Runtime exposure boundary - AI systems may receive PII, PCI, or PHI when they process data:
   - Receive user prompts
   - Retrieve enterprise records to answer questions
   - Assemble responses that may include regulated information
2. Persistent exposure boundary - sensitive data can persist where AI stores or propagates data after processing:
   - Training and fine-tuning datasets
   - Vector databases and embeddings
   - Prompt logs and telemetry systems
   - AI agents and orchestration workflows
   - Human review environments
   - Generated outputs reused across business systems
   - External AI services

CipherTrust Tokenization solutions are architectural safeguards that replace sensitive data with tokens before the data enters AI workflows. Using Tokenization, AI systems receive context, structure and referential integrity without receiving regulated data as cleartext.

CipherTrust Tokenization Solutions

Protecting Runtime and Persistent Exposure Boundaries

CipherTrust Tokenization solutions include three interoperable Connectors to start protecting sensitive data as soon as the data enters your organization at the application layer. The three interoperable Connectors share policies, key management and tokens.

Interoperability ensures that protection applied at the application layer remains consistent across APIs, databases, analytics systems, and AI platforms.

The single-policy model ensures consistent protection across applications, APIs, data pipelines, and AI platforms, eliminating fragmented controls and inconsistent enforcement.

- CipherTrust Application Data Protection (CADP SDK)
- CipherTrust RESTful Data Protection (CRDP RESTful API)
- CipherTrust Data Protection Gateway (DPG)

CADP SDK

CADP integrates directly into applications and AI-enabled services via SDKs.

Developers tokenize sensitive fields before prompts are processed or records are retrieved. Detokenization is tightly controlled and policy-driven.

Fine-grained, in-code control enables high-performance AI workloads without sacrificing governance.

CRDP RESTful API

CRDP provides RESTful tokenization and detokenization for API-centric architectures.

It secures microservices, serverless workflows, and AI-driven applications without requiring embedded SDKs.

Organizations maintain centralized policy control while enabling flexible, cloud-native AI architectures.

DPG API Gateway

DPG operates as a proxy to transparently tokenize or detokenize data flowing between RESTful services.

It protects legacy databases, retrieval pipelines, and RAG architectures without application rewrites.

Security teams can extend protection to AI workloads without slowing modernization efforts.

Downstream Confidence

CipherTrust Tokenization solutions provide downstream confidence because they generate and manage tokens that remain consistent across applications, APIs, data pipelines and AI platforms. Consistency empowers:

- AI-generated outputs to safely flow into CRM systems
- Analytics platforms to join datasets without reintroducing PII
- Fraud systems to detect patterns without raw identifiers
- BI systems to operate without expanding compliance scope

Security, privacy, and innovation move forward together without tension.

Thales Tokenization Solution Benefits

- Eliminate sensitive data from AI systems across hybrid, multi-cloud and on-prem environments
- Reduce PCI DSS 4.0 and privacy compliance scope
- Preserve analytics, RAG, and ML functionality
- Eliminate fragmented controls and inconsistent enforcement
- Ensure that protection applied at the application layer remains consistent across APIs, databases, analytics systems, and AI platforms

Core Use Cases

- Secure AI Chatbots and Copilots
- Safe Retrieval-Augmented Generation (RAG)
- AI Training and Fine-Tuning
- Protect Vector Databases, SQL Databases and Data Lakehouses

Each use case is dependent upon a single objective: sensitive data never enters AI systems as cleartext.

Conclusion

CipherTrust Tokenization solutions enable the use of AI workflows and BI analytics to run unchanged without using sensitive data. Removing sensitive data from the system ensures it can never enter AI systems as cleartext and never propagate beyond intended boundaries. It simplifies compliance and deployment, lowers audit costs, and strengthens your ability to respond confidently to regulators, customers, and the board.

With unified, policy-driven tokenization across application, API, and data layers, AI can learn patterns without learning your sensitive data.

About Thales

Thales is a global leader in cybersecurity, helping businesses, governments, and the most trusted organizations in the world protect critical applications, sensitive data, identities, and software anywhere, at scale — with the highest ROI. With more than 30,000 customers, including 58% of the Fortune Global 500, our solutions are deployed in 148 countries around the world. Through our innovative services and integrated platforms, Thales helps customers achieve better visibility of risks, defend against cyber threats, close compliance gaps, and deliver trusted digital experiences for billions of consumers every day.