

About

Tonic.ai helps organizations create safe, high-quality test data by synthesizing real data while protecting sensitive information like PII and PHI. Our solutions handle structured and unstructured data, and can generate fully synthetic data-from-scratch.

Trusted by Fortune 500 companies and companies across the world, Tonic's customers use us to de-identify, synthesize, and redact PHI and PII across their enterprises to support rapid product development while maintaining compliance for HIPAA, GDPR, and other critical measures.

Tonic Structural

Tonic Structural is a synthetic data generation tool designed to generate safe, realistic datasets for use in staging environments, local development, and testing workflows. It specializes in transforming sensitive structured and semi-structured data, such as PII and PHI, into de-identified or fully synthetic datasets that maintain the statistical properties and relationships of the original data. ([Tonic Structural User Guide](#))

Key Features of Tonic Structural:

- [Data Transformation](#): Applies various data transformation techniques, including masking, subsetting, and synthesis, to protect sensitive information while preserving data utility.
- [Subsetting](#): Allows users to create smaller, representative datasets that maintain referential integrity, facilitating efficient testing and development.
- [Integration Capabilities](#): Supports integration with a wide range of data sources, including relational databases like MySQL, PostgreSQL, SQL Server, and Oracle, as well as platforms like Snowflake and Databricks.
- [Security and Compliance](#): Offers features like role-based access control (RBAC), single sign-on (SSO) integration, and audit logging to help organizations meet



compliance requirements such as HIPAA, GDPR, and CCPA.

Tonic Structural Use Cases:

Category	Use Cases
Software Development	Development Staging QA Local Development Front & Back-End Engineering Performance Testing UAT
Team Collaboration	Enable Offshore Resources Bug Fixes Edge Cases
Security & Compliance	Data Governance Compliance with: HIPAA CCPA GDPR SOC 2
Data Utilization	Reporting & Analytics Model Training Subsetting Database Migrations
Product Operations	Demo Environments Sandbox Testing

By leveraging Tonic Structural, organizations can accelerate development cycles, reduce the risk of exposing sensitive data, and ensure compliance with data protection regulations. It's particularly beneficial for teams looking to enable safe data sharing,

streamline testing processes, and maintain high standards of data privacy. ([Integrating Tonic Structural with your existing tech stack](#))

Tonic Textual

Tonic Textual is designed to synthesize or redact sensitive data within unstructured data sources, like emails, transcripts, medical notes, customer communications, and other document-based content that may contain sensitive information while preserving the utility of the data for downstream use like testing, analytics, and AI/ML training. ([Tonic Textual User Guide](#))

Key Capabilities of Tonic Textual:

- [Named Entity Recognition \(NER\)](#):
Uses proprietary, multilingual machine learning models to accurately detect sensitive entities (e.g., names, dates, locations, medical terms, etc.) across unstructured text.
- [De-Identification](#):
Automatically redacts, masks, or replaces sensitive information to ensure data privacy while maintaining the structure and usefulness of the content.
- **Synthetic Text Generation**:
Replaces identified sensitive entities with realistic but fake alternatives, allowing safe use in AI/ML model training, testing, or sharing.
- [Output Formatting](#):
Converts various unstructured data sources into a standardized, structured output format (such as Markdown or JSON), making it easier to process for downstream use, including LLM pipelines.
- [Multilingual Support](#):
Capable of detecting and de-identifying entities in multiple languages, making it ideal for global datasets.
- [Audio Synthesis](#):
Textual's audio synthesis enables redacting, synthesizing, or "bleeping" our



sensitive information in an audio file, such as names, addresses, or other PII, all while preserving the original structure and tone of the audio. ([Transcribe and redact audio files in Textual](#))

Common Tonic Textual Use Cases:

Category	Use Case	Description
GenAI / LLMs	LLM Development	Preparing compliant, high-quality training data for LLMs
	GenAI Model Training	Using anonymized or synthetic text data to safely train generative models.
	ETL for LLMs	Extracting, transforming, and anonymizing text data before model ingestion.
	LLM Fine-Tuning	Customizing pre-trained models with secure, de-identified domain-specific data.
Security & Compliance	Offshore Resources	Enabling global dev teams with safe, compliant access to datasets.
	Data Governance	Ensuring auditability and traceability of how text data is processed prior to model ingestion.

	HIPAA / GDPR / SOC 2	Meeting regulatory requirements by redacting or synthesizing PII/PHI.
Development & QA	Local Development	Providing engineers with usable yet secure data for testing and dev.
	Reporting & Analytics	Supplying safe data for dashboards, NLP, or BI tools.
	Model Training	Training custom models using redacted or synthesized data.
	Other	Includes customer support training, chatbot design, prompt testing, etc.

By leveraging Tonic Textual, users are enabled with safe, scalable unstructured data by removing or synthesizing sensitive information without sacrificing utility.

Tonic Fabricate

Tonic Fabricate is designed to generate high-fidelity, production-like synthetic data across structured databases and unstructured datasets. It enables developers, testers, and data scientists to work with safe, realistic data that preserves statistical properties, constraints, and business logic. (See the [Tonic Fabricate Product Documentation](#))

Key Capabilities of Tonic Fabricate

- [Schema-Aware Synthesis](#): Fabricate automatically understands and preserves table schemas, primary/foreign key relationships, and constraints when generating synthetic data.
- [Statistical Fidelity](#): The system learns distributions, correlations, and data patterns to generate statistically realistic data that mirrors production environments.
- [Constraint Preservation](#): Maintains business rules, uniqueness, referential integrity, and custom-defined logic so synthetic datasets are usable in downstream applications without manual cleanup.
- [Custom Generators](#): Developers can define custom generation logic for columns, tables, or datasets, allowing tight control over the realism and behavior of the data.
- [Scalability](#): Capable of generating datasets ranging from small test samples to enterprise-scale.

Common Tonic Fabricate Use Cases

Category	Use Case	Description
GenAI / LLMs	Training Data Generation	Producing large-scale synthetic tabular data to train models without exposing sensitive information.
	Fine-Tuning	Creating domain-specific synthetic datasets for secure model fine-tuning.

Development & QA	Local Development	Providing developers realistic, non-sensitive databases for building and testing.
	CI/CD Testing	Automatically generating synthetic datasets for integration and regression testing pipelines.
	Load & Performance Testing	Stress-testing systems with scalable, production-like data.
Security & Compliance	Offshore Development	Equipping offshore teams with safe, compliant data.
	Data Governance	Supporting traceability and auditability of how synthetic data is generated and used.
	Regulatory Compliance	Ensuring synthetic datasets support HIPAA, GDPR, and SOC 2 compliance.
Analytics & BI	Reporting & Dashboard Testing	Feeding BI tools with synthetic data that behaves like production without revealing real customer information.
Other	Data Sharing	Supplying realistic, safe datasets to partners, vendors, or contractors.



By leveraging Tonic Fabricate, teams can safely test, develop, and analyze using realistic synthetic data that mirrors production without putting sensitive data at risk.

Tonic.ai Resources

- [Document Center](#)
- [Tutorial Videos](#)
- [Data Transformation Guides](#)
- [Case Studies](#)
- [Webinars](#)
- [Blog](#)
- [FAQ](#)
- [Trust Center](#)

