

Tonic.ai platform overview

About Tonic.ai

Tonic.ai frees developers to build with safe, high-fidelity synthetic data that accelerates software and AI innovation while protecting data privacy. Through agentic solutions for data synthesis, de-identification, and subsetting, Tonic gives engineering teams on-demand access to realistic data and simulated worlds for development, testing, and AI model and agent training.

Three products do the work, each adopted independently and each interoperating with the others:

- **Tonic Structural** transforms sensitive production data into safe, high-fidelity test data. AI-powered test data management.
- **Tonic Textual** detects sensitive information in unstructured data and redacts or synthesizes it for secure AI development.
- **Tonic Fabricate** generates referentially intact synthetic data from a plain-language prompt, with no production data required.

Customers use them to keep PII and PHI out of development, testing, and AI pipelines while supporting compliance with HIPAA, GDPR, and other data protection obligations.

Why this matters to your business

- **Data access sets the pace of your roadmap.** Test data arrives as a ticket, staging goes stale, and AI projects wait on legal review. The delay shows up as release frequency and escaped defects, not as a data problem.
- **Safe data removes the wait.** Your teams get data that behaves like production with the sensitive content taken out, on demand and without an approval cycle for each request.
- **Compliance becomes an enabler.** When PII and PHI never reach lower environments or AI pipelines, security review is about the work rather than about the data.
- **The scope is both structured and unstructured.** Databases, documents, transcripts, and audio, under one set of controls.



Using de-identified data for AI and agent development

AI agents are only useful when they act on your actual business data: customer records, transactions, support tickets, contracts, and clinical notes. That is also the data you cannot hand out. Most organizations respond by giving a small central AI team access and putting everyone else in a queue, which caps AI output at the size of that team rather than at the number of good ideas in the company.

De-identified data changes that shape:

- **Prototyping goes wide.** With a safe copy of production data in your warehouse, engineers, analysts, and domain experts outside the AI team can build and test agents against realistic data on day one.
- **Prototypes promote without a rewrite.** De-identification preserves structure and referential integrity, so schemas, relationships, and query logic are the same in both places. An agent that works against the safe environment moves to production by repointing it at the production source.
- **Security review narrows.** No sensitive data enters the prototyping environment, so the conversation is about what the agent does rather than about what it might expose.
- **Unstructured data comes with it.** Emails, transcripts, contracts, and notes become available to RAG systems and fine-tuning once sensitive entities are redacted or synthesized.
- **You can test agents before you trust them.** Where no real data exists, simulated environments with known ground truth give you eval suites and regression testing that do not depend on hand-annotating real data.

Where each product fits in AI work:

- **Tonic Structural.** Safe, structurally faithful copies of production databases for teams to build and test agents against.
- **Tonic Textual.** Emails, transcripts, contracts, and clinical notes made safe for RAG, fine-tuning, and model training.
- **Tonic Fabricate.** Simulated environments and graded task sets with known ground truth, for testing, evaluating, and training agents.



Tonic Structural

AI-powered test data management.

Tonic Structural transforms sensitive production data into safe, high-fidelity test data for staging environments, local development, and testing workflows. It de-identifies and synthesizes structured and semi-structured data while preserving the structure and referential integrity your developers need to build and test against. Test data stops being a ticket and becomes self-service. ([Tonic Structural User Guide](#))

Key capabilities of Tonic Structural

- **Agentic configuration.** Configuration is the bulk of the work in any test data project. The built-in [Structural Agent](#) compresses it from hours to minutes, through conversation rather than manual setup. ([What you can ask the Agent to do](#))
- **Sensitivity detection.** A [sensitivity scan](#) identifies columns that contain sensitive information, with custom rules for the types specific to your data.
- **Masking and synthesis.** De-identification and synthesis that maintain your data's structure and preserve referential integrity across complex schemas, so the data stays usable after it is made safe.
- **Patented subsetting.** Targeted, isolated datasets that give each developer their own environment and eliminate test collisions, shrinking large databases to workable size without breaking foreign keys.
- **Native connectors.** Relational databases including MySQL, PostgreSQL, SQL Server, and Oracle, plus Snowflake, Databricks, BigQuery, MongoDB, Salesforce, and file sources. ([Data connector summary](#))
- **Works in your developers' tools.** The [Structural MCP server](#) lets AI clients drive Structural with the same natural-language prompts you use in the Agent.
- **Verification and enterprise controls.** A [Privacy Report](#) documents how each column is protected, alongside role-based access control, SSO integration, and audit logging.

Where teams use Structural

- **Software development.** Development, staging, and QA. Local development. Front and back-end engineering. Performance testing and UAT. Team collaboration and offshore resources. Bug fixes and edge cases.
- **AI and agent development.** Production-like environments for building and testing agents. Broad internal access to realistic data without exposing PII or PHI. Prototypes that promote to production without a rewrite.



- **Security and compliance.** Data governance. Supports HIPAA, CCPA, GDPR, and SOC 2 obligations.
- **Data utilization.** Reporting and analytics. Subsetting. Database migrations.
- **Product operations.** Demo environments. Sandbox testing.

With Structural, your teams accelerate development cycles, reduce the risk of exposing sensitive data, and work from data that stays usable after it has been made safe. ([How Structural fits your existing stack](#))

Tonic Textual

AI-powered unstructured data redaction and synthesis for secure AI development.

Tonic Textual synthesizes or redacts sensitive information in unstructured sources: emails, transcripts, clinical notes, contracts, customer communications, and other document-based content. The data stays useful for testing, analytics, RAG, and AI training. Unstructured data is the largest untapped asset most organizations have, and Textual is what makes it available without manual review that does not scale. ([Tonic Textual User Guide](#))

Key capabilities of Tonic Textual

- **Proprietary NER models.** Purpose-built multilingual models detect sensitive entities in real-world domain text, including clinical notes, legal documents, and transcripts, where generic PII detection underperforms.
- **Redaction and synthesis.** Textual redacts, masks, or replaces sensitive values with realistic synthetic ones, so datasets stay usable for training, testing, and analytics rather than being gutted.
- **Custom entity types.** Train [model-based entity types](#) for the sensitive information specific to your domain.
- **Agentic configuration.** The built-in [Textual Agent](#) lets you explore dataset contents, configure entity handling, and fine-tune synthesis outputs through conversation.
- **Unstructured Data Catalog.** Discover and catalog sensitive unstructured data across sources, organized by entity type, with detection confidence and transformations tracked and natural-language search over the results.
- **Output formatting.** Converts unstructured sources into standardized structured output such as Markdown or JSON, ready for LLM and RAG pipelines.



- **Audio synthesis.** Redact, synthesize, or bleep sensitive information in audio files while preserving the original structure and tone. ([Transcribe and redact an audio file](#))
- **Fits your pipeline.** A Python SDK, an [MCP server](#), and a [Snowflake Native App](#) for redaction inside the warehouse.
- **Auditability.** An [audit log](#) records access, dataset changes, and file activity.

Where teams use Textual

- **GenAI and LLMs.** LLM development and fine-tuning on compliant training data. Safe training of generative models. ETL for LLMs. RAG systems built on internal documents.
- **AI agent development.** Free-text context made safe for agents to read and act on.
- **Security and compliance.** Safe, compliant dataset access for offshore and global teams. Data governance and traceability prior to model ingestion. Helps teams meet HIPAA, GDPR, and SOC 2 obligations.
- **Development and QA.** Usable, secure data for local development. Safe data for dashboards, NLP, and BI. Model training on redacted or synthesized data.
- **Other.** Customer support training, chatbot design, prompt testing.

With Textual, your unstructured data becomes usable in AI development without sacrificing privacy or spending analyst time on manual review.

Tonic Fabricate

Synthetic data generation, as simple as asking for it.

Tonic Fabricate generates and operationalizes referentially intact synthetic data across your entire data ecosystem. Connect to live data sources to model from real databases, generate from scratch, or combine the two. No production data is required, which means no dependency on the access approvals that block most teams. ([Tonic Fabricate documentation](#))

Key capabilities of Tonic Fabricate

- **Agentic generation.** Describe what you need in plain language and the [Data Agent](#) generates it. This is the state-of-the-art approach to synthetic data, and it replaces scripting and rule-writing with a conversation. ([Prompting the Data Agent](#))
- **Plan mode for complex schemas.** Fabricate drafts a strategic generation plan that gives you control over every step before anything is generated.



- **Validation Agent.** Enable it to review the generated code and data and refine what the Data Agent produces, keeping quality high even from an imprecise prompt. ([Using the validation agent](#))
- **Referential integrity across the ecosystem.** Multiple databases and files, with relationships and constraints maintained throughout.
- **Structured and unstructured output.** Generate PDFs, Word files, presentations, email messages, and images populated from the same database. ([Generating unstructured files from a database](#))
- **Works where developers work.** The [Fabricate MCP server](#) brings generation into clients like Claude, Cursor, and VS Code. Request data in plain language and pull finished datasets back into your environment.
- **Operationalization.** [Automated workflows](#) and [mock API endpoints](#) that slot into your pipelines, so you get a repeatable data supply rather than a one-off dataset.
- **Simulated environments for agent work.** From a simple prompt, Fabricate simulates a complete company's activity over a defined timeline, including emails, Slack messages, CRM records, and calendar events, across any tool stack, with temporal integrity and cross-dimensional consistency that algorithmic generation cannot match. Because Fabricate knows what it generated, it can design hierarchies of verifiable tasks of controllable difficulty on top of that environment, from single-hop lookups to multi-hop reasoning chains, each with known ground truth. The same environments work as reinforcement learning training grounds for teams training as well as evaluating.
- **Scale.** From small test samples to enterprise-scale datasets.

Where teams use Fabricate

- **Agent testing, evals, and RL.** Simulated environments with known ground truth. Graded task hierarchies of controllable difficulty. Regression testing. RL training environments.
- **Development and QA.** Realistic, non-sensitive databases for local development. CI/CD and integration testing. Load and performance testing. API and microservices integration testing.
- **GenAI and LLMs.** Large-scale synthetic training data. Domain-specific datasets for fine-tuning.
- **Security and compliance.** Safe, compliant data for offshore teams. Traceability of how synthetic data is generated and used. Supports HIPAA, GDPR, and SOC 2 obligations with no production-data dependency.



- **Analytics and BI.** Reporting and dashboard testing on data that behaves like production.
- **Other.** Safe, realistic datasets for partners, vendors, and contractors.

With Fabricate, developers and AI engineers get the data they need without waiting for access to data they were never going to be given.

How the products work together

- **Structural and Fabricate: scale up de-identified test data.** De-identify production data with Structural, then connect Fabricate to that output to generate additional data modeled on it. You scale volume up without reintroducing sensitive information, which is what performance and load testing needs and what production alone cannot safely provide.
- **Textual and Fabricate: augment unstructured data.** De-identify unstructured datasets with Textual, then point Fabricate at them as inspiration to generate new unstructured data modeled on the originals. Data augmentation for cases where limited training data is the constraint.
- **Structural and Textual: one standard across both data types.** Structured databases and free text handled under the same governance, so nothing falls between the two.

Security and compliance

- Role-based access control, SSO, and audit logging across the products, designed to support obligations under HIPAA, GDPR, CCPA, and SOC 2.
- Tonic.ai maintains a SOC 2 Type II report, validated by an independent auditor.
- HIPAA attestations are available through Tonic.ai's Trust Documents portal.
- Tonic.ai conducts annual application penetration tests on each product, plus annual third-party static code analysis.
- During regular job duties, Tonic.ai staff can never see data processed or stored by Tonic.ai applications.
- Structural and Fabricate offer EU instances in Frankfurt. For Textual, self-hosted deployment is the route to EU data residency.

Policies, SOC 2 reports, penetration test reports, and the subprocessor list are available through the [Tonic.ai Trust Center](#).



Get started

- **Try Fabricate today.** There is a free tier at fabricate.tonic.ai.
- For Structural or Textual, the next step is a technical conversation with our solutions team about your data and your environment.
- Contact: hello@tonic.ai

