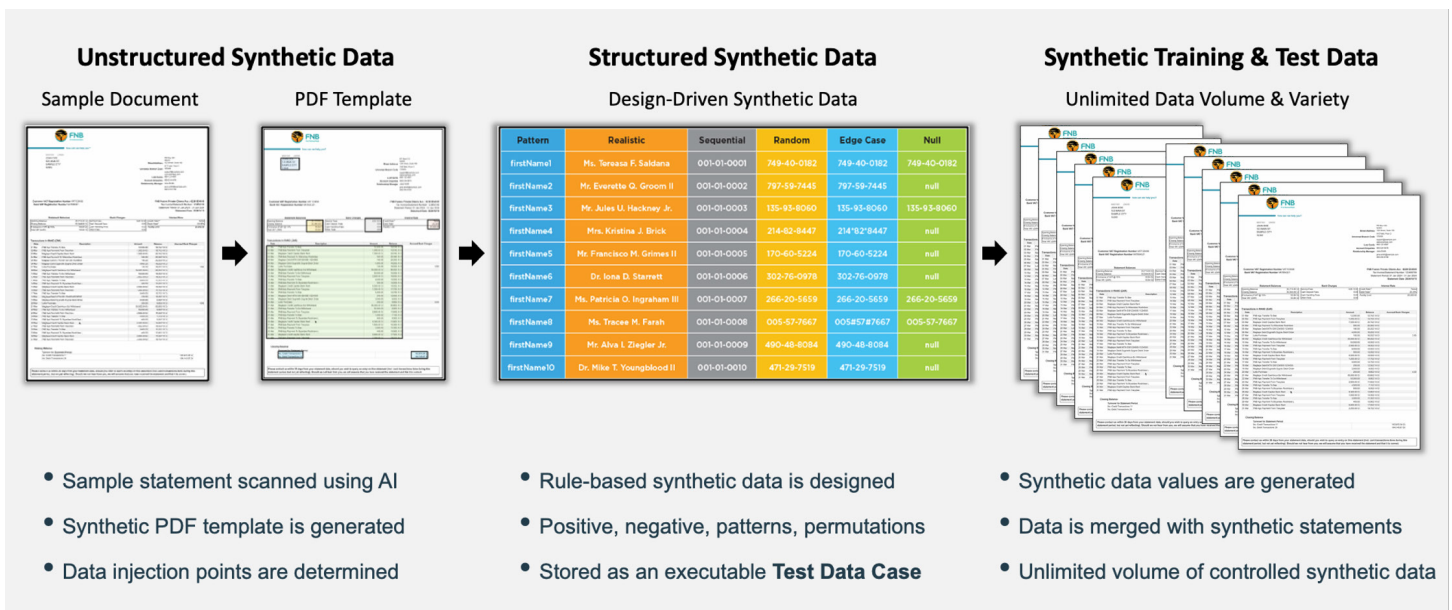


SOLUTIONS GUIDE

UNSTRUCTURED DATA ACCELERATOR (UDA)

UDA Capabilities Overview

The GenRocket Unstructured Data Accelerator (UDA) is a breakthrough enhancement for the GenRocket Synthetic Data generation platform. **UDA allows unstructured data or media to be combined with structured synthetic data that is controlled and conditioned to meet any training or test data requirement.** Now enterprises can quickly and easily generate unstructured data in the form of documents, ID cards, audio clips, or conversational text and blend it with structured data for testing existing OCR and document processing systems or training new and advanced AI-driven applications.



For organizations in regulated industries like financial services and healthcare, UDA makes it possible to safely test, train, and automate document-heavy workflows without accessing or exposing sensitive customer or patient data.

Design-Driven Structured Synthetic Data

One of the key capabilities of UDA is the ability to generate controlled and conditioned structured (tabular) synthetic data to provide complete coverage of all training and test data scenarios. This capability, called **Design-Driven Data**, allows control over the volume, variety, and format of structured synthetic data that is generated on-demand and in real-time. By leveraging GenRocket's more than 750 intelligent data generators, synthetic data can be produced with any or all of the following conditions.

- Positive data
- Negative or null data
- Edge cases and boundary values
- Patterns and permutations
- Statistical distribution profiles
- Amplification of key signals
- Data bias mitigation
- Full referential integrity
- 100% privacy law compliance
- Scalability to billions of rows, records, or documents

Flexible Unstructured Synthetic Data

Equally powerful is the ability to combine structured data with virtually unlimited use cases for unstructured synthetic data in the form of PDF documents, images, text and event streams, and audio files. If physical documents are being digitized for critical processing functions such as check cashing or insurance claims processing, it's essential these documents be captured and converted accurately.

In the case of check cashing, applications can be trained with a variety of positive and negative and edge cases in terms of image recognition. Below is an example of synthetic bank checks that were designed and generated for variety of challenging use cases to train image processing and OCR systems for accuracy and predictability.



Bank checks simulating positive and negative image capture scenarios

Typical scenarios might include images of crumpled paper, torn pages, coffee stains and ink blots that compromise the form data. Similarly, a variety of handwriting styles, legibility, and ink colors may be encountered by these systems. GenRocket's UDA solution has the ability to simulate all of these positive and negative scenarios at scale.

HEALTH INSURANCE CLAIM FORM

APPROVED BY NATIONAL UNIFORM CLAIM COMMITTEE (01/03/2002)

PERSONAL INFORMATION												PATIENT AND INSURED INFORMATION											
1. NAME a. FIRST NAME: <u>John</u> b. LAST NAME: <u>Smith</u> c. SUFFIX: <u>III</u>												2. DATE OF BIRTH: <u>06-09</u> 3. SEX: <u>M</u> 4. SOCIAL SECURITY NUMBER: <u>123-45-6789</u>											
5. ADDRESS a. STREET: <u>123 Main St</u> b. CITY: <u>Anytown</u> c. STATE: <u>CA</u> d. ZIP: <u>90210</u>												6. EMPLOYER'S NAME AND ADDRESS a. NAME: <u>ABC Corp</u> b. ADDRESS: <u>456 Business Ave</u> c. CITY: <u>Anytown</u> d. STATE: <u>CA</u> e. ZIP: <u>90210</u>											
7. POLICY INFORMATION a. POLICY NUMBER: <u>123456789</u> b. POLICY TYPE: <u>Individual</u> c. POLICY EFFECTIVE DATE: <u>01-01-01</u> d. POLICY EXPIRATION DATE: <u>12-31-05</u>												8. CLAIM INFORMATION a. CLAIM NUMBER: <u>123456789</u> b. CLAIM DATE: <u>06-09</u> c. CLAIM TYPE: <u>Medical</u> d. CLAIM STATUS: <u>Submitted</u>											
9. PHYSICIAN INFORMATION a. PHYSICIAN NAME: <u>Dr. John Doe</u> b. PHYSICIAN ADDRESS: <u>123 Medical Dr</u> c. PHYSICIAN CITY: <u>Anytown</u> d. PHYSICIAN STATE: <u>CA</u> e. PHYSICIAN ZIP: <u>90210</u>												10. INSURANCE INFORMATION a. INSURANCE COMPANY: <u>ABC Insurance Co</u> b. INSURANCE ADDRESS: <u>456 Insurance Ave</u> c. INSURANCE CITY: <u>Anytown</u> d. INSURANCE STATE: <u>CA</u> e. INSURANCE ZIP: <u>90210</u>											
11. CLAIM DETAILS a. CLAIM TYPE: <u>Medical</u> b. CLAIM DATE: <u>06-09</u> c. CLAIM STATUS: <u>Submitted</u> d. CLAIM AMOUNT: <u>\$1000.00</u>												12. OTHER INFORMATION a. OTHER CLAIM NUMBER: <u>123456789</u> b. OTHER CLAIM DATE: <u>06-09</u> c. OTHER CLAIM STATUS: <u>Submitted</u> d. OTHER CLAIM AMOUNT: <u>\$1000.00</u>											
13. SIGNATURE a. CLAIMANT SIGNATURE: <u>John Smith</u> b. CLAIMANT ADDRESS: <u>123 Main St</u> c. CLAIMANT CITY: <u>Anytown</u> d. CLAIMANT STATE: <u>CA</u> e. CLAIMANT ZIP: <u>90210</u>												14. PHYSICIAN SIGNATURE a. PHYSICIAN SIGNATURE: <u>Dr. John Doe</u> b. PHYSICIAN ADDRESS: <u>123 Medical Dr</u> c. PHYSICIAN CITY: <u>Anytown</u> d. PHYSICIAN STATE: <u>CA</u> e. PHYSICIAN ZIP: <u>90210</u>											
15. CLAIM HISTORY a. CLAIM TYPE: <u>Medical</u> b. CLAIM DATE: <u>06-09</u> c. CLAIM STATUS: <u>Submitted</u> d. CLAIM AMOUNT: <u>\$1000.00</u>												16. OTHER INFORMATION a. OTHER CLAIM NUMBER: <u>123456789</u> b. OTHER CLAIM DATE: <u>06-09</u> c. OTHER CLAIM STATUS: <u>Submitted</u> d. OTHER CLAIM AMOUNT: <u>\$1000.00</u>											
17. CLAIM DETAILS a. CLAIM TYPE: <u>Medical</u> b. CLAIM DATE: <u>06-09</u> c. CLAIM STATUS: <u>Submitted</u> d. CLAIM AMOUNT: <u>\$1000.00</u>												18. OTHER INFORMATION a. OTHER CLAIM NUMBER: <u>123456789</u> b. OTHER CLAIM DATE: <u>06-09</u> c. OTHER CLAIM STATUS: <u>Submitted</u> d. OTHER CLAIM AMOUNT: <u>\$1000.00</u>											
19. CLAIM HISTORY a. CLAIM TYPE: <u>Medical</u> b. CLAIM DATE: <u>06-09</u> c. CLAIM STATUS: <u>Submitted</u> d. CLAIM AMOUNT: <u>\$1000.00</u>												20. OTHER INFORMATION a. OTHER CLAIM NUMBER: <u>123456789</u> b. OTHER CLAIM DATE: <u>06-09</u> c. OTHER CLAIM STATUS: <u>Submitted</u> d. OTHER CLAIM AMOUNT: <u>\$1000.00</u>											
21. CLAIM DETAILS a. CLAIM TYPE: <u>Medical</u> b. CLAIM DATE: <u>06-09</u> c. CLAIM STATUS: <u>Submitted</u> d. CLAIM AMOUNT: <u>\$1000.00</u>												22. OTHER INFORMATION a. OTHER CLAIM NUMBER: <u>123456789</u> b. OTHER CLAIM DATE: <u>06-09</u> c. OTHER CLAIM STATUS: <u>Submitted</u> d. OTHER CLAIM AMOUNT: <u>\$1000.00</u>											
23. CLAIM HISTORY a. CLAIM TYPE: <u>Medical</u> b. CLAIM DATE: <u>06-09</u> c. CLAIM STATUS: <u>Submitted</u> d. CLAIM AMOUNT: <u>\$1000.00</u>												24. OTHER INFORMATION a. OTHER CLAIM NUMBER: <u>123456789</u> b. OTHER CLAIM DATE: <u>06-09</u> c. OTHER CLAIM STATUS: <u>Submitted</u> d. OTHER CLAIM AMOUNT: <u>\$1000.00</u>											
25. CLAIM DETAILS a. CLAIM TYPE: <u>Medical</u> b. CLAIM DATE: <u>06-09</u> c. CLAIM STATUS: <u>Submitted</u> d. CLAIM AMOUNT: <u>\$1000.00</u>												26. OTHER INFORMATION a. OTHER CLAIM NUMBER: <u>123456789</u> b. OTHER CLAIM DATE: <u>06-09</u> c. OTHER CLAIM STATUS: <u>Submitted</u> d. OTHER CLAIM AMOUNT: <u>\$1000.00</u>											
27. CLAIM HISTORY a. CLAIM TYPE: <u>Medical</u> b. CLAIM DATE: <u>06-09</u> c. CLAIM STATUS: <u>Submitted</u> d. CLAIM AMOUNT: <u>\$1000.00</u>												28. OTHER INFORMATION a. OTHER CLAIM NUMBER: <u>123456789</u> b. OTHER CLAIM DATE: <u>06-09</u> c. OTHER CLAIM STATUS: <u>Submitted</u> d. OTHER CLAIM AMOUNT: <u>\$1000.00</u>											
29. CLAIM DETAILS a. CLAIM TYPE: <u>Medical</u> b. CLAIM DATE: <u>06-09</u> c. CLAIM STATUS: <u>Submitted</u> d. CLAIM AMOUNT: <u>\$1000.00</u>												30. OTHER INFORMATION a. OTHER CLAIM NUMBER: <u>123456789</u> b. OTHER CLAIM DATE: <u>06-09</u> c. OTHER CLAIM STATUS: <u>Submitted</u> d. OTHER CLAIM AMOUNT: <u>\$1000.00</u>											
31. CLAIM HISTORY a. CLAIM TYPE: <u>Medical</u> b. CLAIM DATE: <u>06-09</u> c. CLAIM STATUS: <u>Submitted</u> d. CLAIM AMOUNT: <u>\$1000.00</u>												32. OTHER INFORMATION a. OTHER CLAIM NUMBER: <u>123456789</u> b. OTHER CLAIM DATE: <u>06-09</u> c. OTHER CLAIM STATUS: <u>Submitted</u> d. OTHER CLAIM AMOUNT: <u>\$1000.00</u>											
33. CLAIM DETAILS a. CLAIM TYPE: <u>Medical</u> b. CLAIM DATE: <u>06-09</u> c. CLAIM STATUS: <u>Submitted</u> d. CLAIM AMOUNT: <u>\$1000.00</u>												34. OTHER INFORMATION a. OTHER CLAIM NUMBER: <u>123456789</u> b. OTHER CLAIM DATE: <u>06-09</u> c. OTHER CLAIM STATUS: <u>Submitted</u> d. OTHER CLAIM AMOUNT: <u>\$1000.00</u>											
35. CLAIM HISTORY a. CLAIM TYPE: <u>Medical</u> b. CLAIM DATE: <u>06-09</u> c. CLAIM STATUS: <u>Submitted</u> d. CLAIM AMOUNT: <u>\$1000.00</u>												36. OTHER INFORMATION a. OTHER CLAIM NUMBER: <u>123456789</u> b. OTHER CLAIM DATE: <u>06-09</u> c. OTHER CLAIM STATUS: <u>Submitted</u> d. OTHER CLAIM AMOUNT: <u>\$1000.00</u>											
37. CLAIM DETAILS a. CLAIM TYPE: <u>Medical</u> b. CLAIM DATE: <u>06-09</u> c. CLAIM STATUS: <u>Submitted</u> d. CLAIM AMOUNT: <u>\$1000.00</u>												38. OTHER INFORMATION a. OTHER CLAIM NUMBER: <u>123456789</u> b. OTHER CLAIM DATE: <u>06-09</u> c. OTHER CLAIM STATUS: <u>Submitted</u> d. OTHER CLAIM AMOUNT: <u>\$1000.00</u>											
39. CLAIM HISTORY a. CLAIM TYPE: <u>Medical</u> b. CLAIM DATE: <u>06-09</u> c. CLAIM STATUS: <u>Submitted</u> d. CLAIM AMOUNT: <u>\$1000.00</u>												40. OTHER INFORMATION a. OTHER CLAIM NUMBER: <u>123456789</u> b. OTHER CLAIM DATE: <u>06-09</u> c. OTHER CLAIM STATUS: <u>Submitted</u> d. OTHER CLAIM AMOUNT: <u>\$1000.00</u>											
41. CLAIM DETAILS a. CLAIM TYPE: <u>Medical</u> b. CLAIM DATE: <u>06-09</u> c. CLAIM STATUS: <u>Submitted</u> d. CLAIM AMOUNT: <u>\$1000.00</u>												42. OTHER INFORMATION a. OTHER CLAIM NUMBER: <u>123456789</u> b. OTHER CLAIM DATE: <u>06-09</u> c. OTHER CLAIM STATUS: <u>Submitted</u> d. OTHER CLAIM AMOUNT: <u>\$1000.00</u>											
43. CLAIM HISTORY a. CLAIM TYPE: <u>Medical</u> b. CLAIM DATE: <u>06-09</u> c. CLAIM STATUS: <u>Submitted</u> d. CLAIM AMOUNT: <u>\$1000.00</u>												44. OTHER INFORMATION a. OTHER CLAIM NUMBER: <u>123456789</u> b. OTHER CLAIM DATE: <u>06-09</u> c. OTHER CLAIM STATUS: <u>Submitted</u> d. OTHER CLAIM AMOUNT: <u>\$1000.00</u>											
45. CLAIM DETAILS a. CLAIM TYPE: <u>Medical</u> b. CLAIM DATE: <u>06-09</u> c. CLAIM STATUS: <u>Submitted</u> d. CLAIM AMOUNT: <u>\$1000.00</u>												46. OTHER INFORMATION a. OTHER CLAIM NUMBER: <u>123456789</u> b. OTHER CLAIM DATE: <u>06-09</u> c. OTHER CLAIM STATUS: <u>Submitted</u> d. OTHER CLAIM AMOUNT: <u>\$1000.00</u>											
47. CLAIM HISTORY a. CLAIM TYPE																							

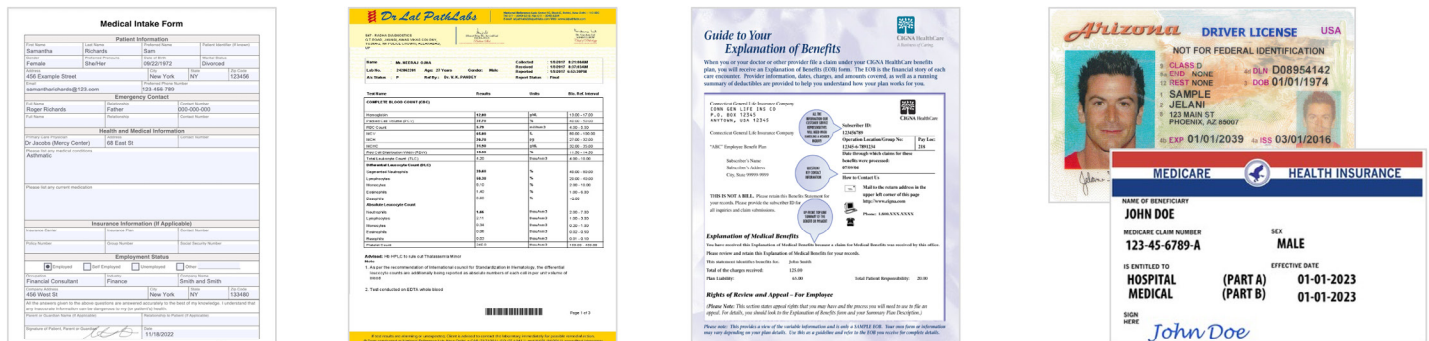
The unique power of UDA is the ability to combine design-driven structured data with flexible unstructured data to produce high quality training and test data. The UDA solution can generate an unlimited volume of synthetic data through the use of executable Test Data Projects that can be categorized, stored, reused, shared, repurposed, and version controlled.

BANK NAME DATE 10/24/11
 PAY TO THE ORDER OF \$
 DOLLARS
 MEMO Authorized Signature
 MICR LINE: ⑈ 436799031 ⑈ 12654321098765432109⑈

- Generate synthetic bank statements, trade confirmations, and wire transfer forms to test transaction workflows without exposing customer data.
- Create synthetic claims forms and loan applications to validate document capture and processing systems.
- Produce regulator-required formats safely in test environments for compliance testing.
- Generate synthetic audio recordings to train call center staff and validate speech analytics tools.

Healthcare

In healthcare, any combination of documents and images can be simulated synthetically. PDF templates can simulate insurance claims and laboratory documents. Scanned identification cards can be generated for testing patient verification and eligibility systems. And audio clips can be used to train and test text-to-speech or speech-to-text systems.



- Create synthetic patient intake forms, billing statements, and insurance ID cards to test EHR and billing workflows in a HIPAA-compliant manner.
- Generate synthetic documents to validate EMR/EHR system workflows without using real patient data.
- Produce synthetic ID cards combined with structured data to support secure identity verification systems.

Customer Challenges Solved with UDA

Compliance risk: Production documents often contain PII/PHI and can't be safely used.

Limited coverage: Real data rarely provides edge cases and negated testing scenarios.

Scaling needs: Training and testing AI/ML systems require massive volumes of data.

Integration gaps: Data must be delivered seamlessly across the DevOps environment.

Benefits Realized

- Accelerates testing and training by providing large, diverse, and realistic datasets.
- Eliminates compliance risk by replacing real documents with synthetic documents.
- Ensures full coverage with the ability to generate unlimited data variations.
- Scales with CI/CD pipelines to support automated and integrated DevOps workflows.
- Boosts AI/ML model accuracy with controlled, balanced, and conditioned data.

How UDA Delivered

UDA is a modular add-on to the GenRocket platform, not a standalone tool. It extends synthetic data into the world of unstructured data with software and professional services that complement the existing GenRocket synthetic data platform. UDA will be configured with the required number of *Test Data Projects* and *Navigator Services* to ensure a successful implementation that meets your unstructured data requirements.

UDA is Now Available for Beta Testing

The Unstructured Data Accelerator is planned for production release in Q4 of 2025. However, the UDA solution is now available for beta testing to organizations that match our beta testing requirements. If your organization is addressing some of the challenges and use cases described in this document, we are ready to explore enrolling your company in our beta program. You can also inquire about the UDA beta program by Email [HERE](#). They will schedule a discovery session with our UDA experts to discuss your specific use case and how GenRocket's UDA can meet your requirements.

