



DATA SHEET

UDA-REDACT

Machine Learning Powered PII Redaction for Unstructured Data

UDA-Redact is GenRocket's machine learning powered unstructured data redaction accelerator designed to permanently remove Personally Identifiable Information (PII) and Protected Health Information (PHI) from enterprise unstructured data including PDFs, scanned forms, images, claims documents, financial records, and identity documents. The platform combines deep learning powered detection, human-in-the-loop review, immutable audit logging, and pixel-level permanent redaction inside a fully offline, Docker-native deployment architecture with zero data egress.

Unlike conventional rules-based or regex-driven approaches, UDA-Redact uses deep learning models trained on the semantic structure of enterprise documents, enabling more accurate detection across structured forms, free-flow text, handwritten regions, and mixed-layout documents.

22222	VOID ☐	a Employee's social security number SSN		For Official Use Only OMB No. 1545-0029			
b Employee identification number (EIN) EIN			1 Wages, tips, other compensation \$112,000.00		2 Federal income tax withheld \$14,160.00		
c Employer's name, address, and ZIP code Gritworks Technology LLC 4125 Westbrook Corporate Center, Suite 240, San Bernardino, CA 92408 California, USA			3 Social security wages \$115,200.00		4 Social security tax withheld \$7,142.40		
			5 Medicare wages and tips \$115,200.00		6 Medicare tax withheld \$1,670.40		
			7 Social security tips		8 Allocated tips		
d Control number CONTROL_NUM			9		10 Dependent care benefits \$0.00		
e Employee's first name and initial FIRST_NAME		Last name LAST_NAME	Self OFF	11 Nonqualified plans \$0.00		12a See instructions for box 12 D \$4,560.00	
FULL_NAME ADDRESS			13 Salaried employee N		Retirement plan Y	Third party sick pay N	12b DD \$7,800.00
			14 Other CA SDI: 1,314.00		12c		
					12d		
f Employee's address and ZIP code							
15 State CA	Employer's state ID number FLOWER_STATE	16 State wages, tips, etc. \$112,000.00	17 State income tax \$3,180.00	18 Local wages, tips, etc. \$0.00	19 Local income tax \$0.00	20 Locality name \$0.00	

Key Features & Capabilities

- Machine learning powered PII and PHI detection automatically identifies sensitive information including Social Security Numbers, patient records, account numbers, addresses, names, and regulated financial and healthcare data.
- Human-in-the-loop operator review and approval ensures every detected region is reviewed by an operator before permanent redaction is applied, enabling governance oversight and continuous model improvement.
- Pixel-level permanent redaction irreversibly removes underlying sensitive information rather than simply overlaying removable masks or annotations.
- Immutable audit logs with SHA-256 chain-of-custody proof create tamper-evident governance records for compliance, legal, and audit teams.
- Governance CSV export provides structured reporting containing detection metadata, confidence scores, timestamps, bounding box coordinates, and file hashes.
- Batch processing supports concurrent processing of high-volume enterprise document workloads for testing, compliance, and AI training preparation.
- 100% offline deployment with zero data egress allows UDA-Redact to operate entirely within the customer's infrastructure without external APIs, cloud uploads, or LLM dependencies.
- Support for PDFs and images today, with additional unstructured formats on the roadmap including text, multimedia, audio, biometrics, and video generation capabilities.

UDA-Redact is available as a standalone accelerator solution for enterprises that need compliant unstructured data redaction capabilities today. It is also extensible to UDA-Generate — GenRocket's enhanced unstructured synthetic data accelerator that is part of the GenRocket Quality Evolution Platform (QEP) for design-driven synthetic data generation.

The combined "Redact → Generate" workflow enables organizations to evolve from simply removing sensitive information to generating controlled, conditioned, rule-based synthetic unstructured data engineered specifically for testing and AI model training objectives. Organizations can preserve real-world document realism while generating unlimited synthetic variants, edge cases, negative scenarios, and scalable AI training corpora without exposing real PII or PHI.

UDA-Redact includes initial platform capacity for redacting up to 5,000 documents through the inclusion of 10 GenRocket Test Data Microprojects. Each Test Data Microproject supports the redaction of up to 500 documents, providing organizations with a scalable and modular capacity model aligned to enterprise growth requirements and project expansion.

Additional Microprojects can be purchased in packs of 10, enabling cost-controlled and predictable scaling of document processing capacity over time. This modular expansion model allows enterprises to begin with a targeted deployment footprint while scaling incrementally as additional testing, compliance, AI, and document-processing initiatives emerge.

Typical enterprise use cases include Application Test Data Management (TDM), AI Model Training, Agentic AI Development, Intelligent Document Processing (IDP), healthcare claims redaction, mortgage and financial document testing, OCR validation, secure document sharing, and synthetic training data preparation.

