WITNESS DATA FACTORY - License and Compliance Framework v1.3

WITNESS DATA FACTORY™

LICENSE AND COMPLIANCE FRAMEWORK v1.3

DOCUMENT TYPE: Commercial License Terms, Tier Definitions, and Compliance Framework **DOCUMENT VERSION:** 1.3 **EFFECTIVE DATE:** June 3, 2026 **ISSUING AUTHORITY:** WITNESS DATA FACTORY™ **PRINCIPAL:** Hector Ortiz, Data Engineering Lead **CONTACT:** WitnessDataFactory@gmail.com **Classification:** Public — May be shared with auditors, regulators, clients, and procurement teams

---

EXECUTIVE SUMMARY

This document defines the commercial license terms, delivery tier definitions, delivery/integrity guarantees, and compliance framework applicable to all datasets purchased or contracted from WITNESS DATA FACTORY™. It is designed for procurement officers, legal counsel, compliance teams, and commercial partners evaluating or licensing WITNESS DATA FACTORY™ synthetic medical datasets. It applies across all distribution channels; the specific channel and commercial terms for any purchase are set out in the applicable offer or order documentation.

SECTION 1: COMMERCIAL TIER DEFINITIONS

1.1 Dataset Tiers

WITNESS DATA FACTORY™ offers the following tiers. The tier defines the dataset size delivered for that purchase:

Tier	Records	Availability
Evaluation	1,000	Free — no purchase required
Research	10,000	Commercial
Professional	50,000	Commercial
Enterprise	250,000	Commercial
Platform	1,000,000	Commercial

All tiers are drawn from a single governed, QA-passed, de-duplicated master dataset and share identical schema and documentation. Smaller tiers are nested subsets of larger ones.

1.2 Delivered Counts

For live data-share / marketplace delivery, each tier delivers an exact row count (e.g., 50,000 means exactly 50,000 records). For per-batch file (ZIP) delivery, the tier denotes the requested batch size and the delivered count after automated QA exclusion is recorded in the batch's QA certificate (`records_delivered`); the delivered count may be slightly lower than the requested size. The authoritative delivered quantity for any purchase is stated in the accompanying documentation (the dataset metadata/checksums for live shares, or the QA certificate for file deliveries). Tier file exports drawn from the governed master (e.g., a marketplace/platform stage export) deliver the exact tier row count, recorded as `record_count` in the tier integrity certificate (`qa_certificate_tier_<tier>.json`) and the delivery manifest. These exact-count exports are distinct from per-batch ZIP generations, whose delivered count (`records_delivered`) may be slightly lower than the requested size.

SECTION 2: DELIVERY AND INTEGRITY

2.1 Delivery Methods

Datasets may be delivered through one or more channels, as specified in the applicable offer:

- **Live data share / marketplace listing** — the buyer queries a shared, governed dataset directly in their own data-platform account; no file download.
- **File package (ZIP)** — the buyer receives dataset files plus machine-readable QA certificate and manifest, and supporting documentation.

The active offer or order documentation specifies the channel and delivery method for each purchase.

2.2 Integrity Verification

- **Live data-share delivery:** integrity is verifiable via reproducible content hashes computed over the exact delivered rows (e.g., an order-independent full-row aggregate hash for all tiers, and a full-row SHA-256 digest for smaller tiers), published in the accompanying dataset documentation. The buyer reproduces the hash on the delivered data and compares.
- **File delivery (ZIP or platform stage export):** integrity is verifiable via (a) file-level cryptographic hashes published in the delivery manifest — MD5 when the file inventory is produced from a data-platform directory table, or SHA-256 for client-built ZIP manifests — and (b) reproducible full-row content hashes (an order-independent aggregate hash for all tiers, plus a full-row SHA-256 ordered digest for tiers up to 250K). The buyer recomputes the file hash and/or reproduces the content hash and compares. Each method's hashes apply only to that delivery form.

Each method's hashes apply only to that delivery form.

2.3 What Is Guaranteed

1. 100% synthetic data — zero PHI, zero PII.
2. Provenance, QA-field definitions, and compliance documentation accompany every delivery.
3. Reproducible integrity verification appropriate to the delivery method.
4. Version-stamped datasets; integrity values are regenerated whenever a dataset is refreshed.

SECTION 3: COMPLIANCE FRAMEWORKS

WITNESS DATA FACTORY™ datasets are designed and documented in alignment with the following compliance frameworks.

3.1 HIPAA-Aligned

WITNESS DATA FACTORY™ synthetic datasets are not Protected Health Information (PHI) as defined under HIPAA 45 CFR §160.103 and §164.514. No Business Associate Agreement (BAA) is required for procurement, receipt, or use of these datasets. Buyers may use these datasets for HIPAA-covered-entity-adjacent work (model training, clinical AI development) without PHI handling obligations.

3.2 GDPR-Positioned as Anonymous Synthetic Data

Datasets do not relate to any identified or identifiable natural person. Generation involves no processing of EU data subjects' personal data. Datasets are positioned as anonymous synthetic data falling outside the material scope of GDPR (Regulation EU 2016/679, Article 2(1)) per Article 4(1) definition of personal data. Academic consensus (IJLIT, January 2026) supports this positioning for properly generated synthetic data.

3.3 EU AI Act Article 10 Supporting Documentation

For buyers training or fine-tuning high-risk AI systems under Regulation EU 2024/1689, WITNESS DATA FACTORY™ datasets and accompanying documentation (this document, the Provenance/Compliance Certificate, and the Technical Proof Addendum) are designed to serve as Article 10 data governance supporting documentation, evidencing dataset relevance, representativeness, and freedom from systematic bias.

3.4 FDA Development-Phase-Ready

In alignment with FDA Guidance (June 2025, "Use of AI in Drug Development"), WITNESS DATA

FACTORY™ synthetic datasets are suitable for:

- AI/ML model pre-training and fine-tuning
- Pre-clinical model validation
- Benchmark construction for model comparison
- Augmenting annotated training sets during development

Synthetic data is not a substitute for real-world clinical validation data required for FDA regulatory clearance or approval of AI-enabled medical devices. Buyers planning regulatory submissions must include real-world data in final clinical validation.

SECTION 4: DELIVERY PACKAGE CONTENTS

Contents depend on the delivery method specified in the offer:

Live data-share / marketplace delivery — shared database objects:

Object	Role
Tier dataset view	The licensed synthetic records
Dataset metadata	Provenance + QA summary
Data dictionary	Field definitions
Provenance / QA-field / compliance documentation	Supporting documentation
Tier list + integrity checksums	Authoritative tier list + integrity hashes

File delivery (ZIP or stage export) — files:

Artifact	Role
Dataset file (JSONL)	Delivered synthetic records
QA / integrity certificate (JSON)	Per-batch ZIP: canonical batch attestation (`qa_certificate_<domain>_req<id>.json`). Stage tier export: tier integrity certificate (`qa_certificate_tier_<tier>.json`) — exact count, all-PASS, reproducible content hashes
Delivery manifest (JSON)	File-level audit trail and hashes
README + VERIFY_INSTRUCTIONS	Human-readable usage + verification steps
verify_package.py	One-command file-integrity verifier (stage exports)
Compliance PDFs	Bundled in ZIP deliveries; available from the provider for stage exports

---

SECTION 5: LICENSE TERMS

5.1 Grant of License

Subject to the terms of this framework and any applicable Order Form or purchase-channel offer terms, WITNESS DATA FACTORY™ grants the buyer a non-exclusive, non-transferable license to use the delivered synthetic datasets for:

- AI/ML model training, fine-tuning, pre-training, and evaluation
- Healthcare software development and testing
- Research and academic publication (with attribution)
- Clinical AI benchmarking and validation (non-final)
- Internal business intelligence and analytics

5.2 Restrictions

The buyer may NOT:

- Re-sell, sub-license, or redistribute raw dataset files to third parties without written authorization
- Claim the synthetic data as real patient data or present it as derived from real clinical records
- Attempt to use the data for final regulatory clinical validation as a substitute for real-world data
- Attempt to reverse-engineer generation methodology using the delivered datasets
- Use the WITNESS DATA FACTORY™ name or brand to imply endorsement of derived products without written authorization

5.3 Disclaimer

WITNESS DATA FACTORY™ synthetic datasets are provided as-is for development and research purposes. They are not a substitute for real-world clinical data in final regulatory submissions. WITNESS DATA FACTORY™ makes no warranty that use of synthetic datasets alone will satisfy any specific regulatory clearance requirement.

SECTION 6: CONTACT AND SUPPORT

- **Email:** WitnessDataFactory@gmail.com
- **Platform:** https://witness-data-factory.onrender.com
- **Compliance inquiries:** WitnessDataFactory@gmail.com — subject line "Compliance Inquiry"
