ELYREON DATASET DOCTOR

Dataset Verification & Quality Audit Report • Job ID: **demo_02_privacy_pii_audit**

1. Executive Summary & Record Statistics

Metric	Value	Metric	Value
Total Raw Records	7	Valid Cleaned Records	7
Modified Records	2	Rejected Records	0
Exact Duplicates	0	Near Duplicates	0
Syntax Errors	0	PII Incidents Detected	2

2. Token Distribution & Length Characteristics

Total Tokens	Min Tokens	Mean Tokens	Median Tokens	95th Percentile
203	21	29.0	29	35

4. Verification Methodology & Scientific Integrity

Validation applies strict deterministic parsing, character-level Unicode NFC normalization, 64-bit SimHash Hamming-distance near-duplicate detection, and RFC-compliant PII scanning. **Notice:** ELYREON Dataset Doctor does not fabricate synthetic quality improvements or guarantee downstream model benchmark performance without empirical training runs.

5. Cryptographic Manifest Checksums

Artifact	SHA-256 Checksum
Input Original File	`762040375267d4e4afe7a4f2d8270df3e9fd1e7a7b897eb37cf1bead6db35874`
Cleaned Dataset	`b0051c4ab850796c4a5005f54a2670fd605b4ca22ed4ff4679c1f14aae4857fc`
Rejected Records	`e3b0c44298fc1c149afbf4c8996fb92427ae41e4649b934ca495991b7852b855`
Change Log	`6f724f02900636db5e3619400f216712494b48916dd809c6a121413299c6adb8`