

Petabyte-Scale eDiscovery with Full PII Protection: A Fortune 100 Financial



Venio's Impact

120+TB → 34TB

72% data reduction
after deduplication

6 Hours

first-look analytics
from initial data drop

95% Precision

PII detection for sensitive
information

2.1x Efficiency

pReview Efficiency faster,
smarter document review

About the firm

A Fortune 100 global bank came under the spotlight of a major regulatory investigation, with custodians spread across North America and EMEA. The data footprint was vast and diverse, spanning Microsoft 365, legacy PSTs, mobile device extractions, and shared network drives.

The bank's mandate was clear and urgent: process massive volumes of data at speed, deliver early insights to outside counsel, and enforce uncompromising controls around personally identifiable information (PII).

The Challenges

- **Overwhelming scale:** 120+ terabytes of data across 38 custodians, complicated by mixed encodings and encrypted archives.
- **Strict privacy demands:** Personally identifiable information (SSNs, account numbers, addresses) required precise detection, redaction, and full auditability.
- **Race against time:** Rolling collections with daily deltas meant outside counsel expected same-day analytics, leaving no room for delays.

How Venio Solved it

Collection & Staging

From the moment data entered the system, Venio ensured defensibility and control. Every intake was tracked with full chain-of-custody, while checksum validation and detailed inventories guaranteed data integrity. Secure staging to AWS S3/FSx allowed the team to manage massive volumes without compromising accessibility or security.

Processing Pipeline

Venio's high-throughput processing pipeline tackled the complexity head-on. Text extraction and OCR handled image-heavy emails and scanned documents, while metadata normalization - covering time zones, message IDs, and conversation threading kept communication threads coherent.

Near-duplicate detection and deduplication, both at the custodian and global levels, significantly reduced the reviewable volume. Language detection and tokenization further enhanced the accuracy of searches and analytics, ensuring no signal was missed.

PII Detection & Early Case Assessment (ECA)

Protecting sensitive data was a top priority. Venio combined built-in PII detection with integrated AI tagging to identify sensitive content such as SSNs and account numbers. These PII filters surfaced in the Early Case Assessment workspace, allowing teams to triage and quarantine sensitive documents before they reached reviewers.

Review

Venio's Review interface (VOD) provided a streamlined experience for legal teams. Reviewers leveraged faceted search, saved searches, and custom queues, while Continuous Active Learning (CAL) prioritized documents most likely to be relevant.

Whereas pattern-based and manual redactions, privilege tagging, annotations, and quality control dashboards ensured thorough, defensible review at scale.

Production

Productions were flexible, accurate, and regulator-ready. Venio generated output with Bates numbering, branding, and multiple formats, including TIFF, JPEG+Text, and PDF. Load files were compatible with Relativity, Concordance, and other platforms, while privilege logs and PII-safe exports ensured compliance and privacy at every stage.



Measurable Results

Metric	Value	Notes
Data Reduction	120 TB → 34 TB	Achieved 72% reduction through custodian-level and global deduplication, & system file suppression.
First-Look Analytics	6 Hours	From initial data drop to actionable insights for counsel.
Daily Delta Turnaround	4 Hours	Continuous processing of new data batches without disrupting ongoing review.
PII Detection Precision	>95%	SSN pattern detection accuracy after reviewer feedback loop integration.
Review Efficiency	2.1× Improvement	CAL-driven prioritization increased reviewer throughput.
Privilege Error Rate	<0.5%	After QC sampling and validation workflows.
Production Acceptance	100% First Pass	All productions accepted by regulators without revision requests.
Data Incidents	Zero	No PII breaches, data leaks, or security incidents throughout the engagement.

Results

The impact of Venio's approach was immediate and measurable. From an initial dataset of 120 terabytes, deduplication and suppression of system files reduced the volume to just 34 terabytes - a 72% reduction that dramatically cut the reviewable workload.

First-look analytics were available within 6 hours of the initial data drop, and rolling daily deltas were processed and ready for counsel in just 4 hours, keeping pace with the fast-moving investigation.

PII detection exceeded expectations, surfacing sensitive information with over 95% precision for Social Security numbers (SSN) after a reviewer feedback loop. Continuous Active Learning (CAL) optimized review, doubling throughput, while privilege error rates dropped below 0.5%, ensuring defensible and accurate results.

Perhaps most importantly, all productions were accepted by regulators on the first pass, eliminating rework and demonstrating the defensibility of the workflow. By combining speed, accuracy, and privacy controls, Venio enabled the bank to tackle a massive, complex investigation efficiently and confidently.

Why Did it Work?

Venio's horizontal processing enabled rapid handling of massive data volumes, while robust PII governance ensured sensitive information was always protected. Integrated Early Case Assessment (ECA) provided counsel with early insights, allowing faster, informed decisions.

The combination of speed, privacy, and visibility made the workflow both efficient and fully defensible.

Key Takeaways

- **Speed and scale don't compromise security:** Robust architecture allows 120+ TB to be processed quickly while keeping PII fully protected.
- **PII detection still needs human validation:** Even >95% precision requires reviewer oversight to ensure defensibility.
- **Daily deltas demand automation:** Manual processes can't sustain 4-hour turnaround cycles.
- **Early Case Assessment (ECA) is essential:** Counsel needs actionable analytics before committing to a full review.
- **Continuous Active Learning (CAL) requires trust:** Efficiency gains only occur when reviewers rely on AI-driven document ranking.

About Venio Systems

At Venio Systems, we are dedicated to working with our trusted partners to bring the latest legal technology innovations to law firms, agencies, and corporations. We combine advanced technology with practical design to deliver smarter eDiscovery. Our all-in-one platform helps organizations streamline workflows, reduce costs, and maintain defensibility at every stage of the EDRM.

Why Venio is Different

Traditional Tools

- Multiple tools, fragmented workflows
- Error-prone manual processes
- Limited reporting visibility
- High per-user licensing costs
- Difficult to scale with modern data types

Venio Systems

- Unified end-to-end platform
- Automated, AI-driven workflows
- Real-time dashboards & analytics
- Flexible, cost-efficient pricing
- Scalable, future-ready architecture



Ready to Rethink eDiscovery?

Book a Demo Today!



VENIO



Venio
Review



Venio
Legal Hold



Venio ECA



Venio
AI Review



Venio
Production