eFraud Investigator

Methodology for Financial Data Extraction & Verification

Technical Memorandum • Forensic Accountants, Fraud Investigators, Legal Counsel & the Courts

To: Forensic Accountants, Fraud Investigators, Legal Counsel, and the Courts
From: eFraud Services Inc., developers of eFraud Investigator
Subject: Methodology for digital financial-data extraction and verification
Version: 1.0 • June 2026

> **ℹ Purpose of this memorandum**
>
> This memorandum describes how eFraud Investigator extracts data from financial records and how that data is verified. While the platform uses AI technology (Google Gemini) for character and layout recognition, it is engineered to function as a constrained transcription tool — not a generative or interpretive one. The intended reader is the qualified professional who must describe the basis of an analysis to a client, an opposing party, or a court.

1. Purpose and Scope

eFraud Investigator is a transcription and organization tool. It captures financial data from source documents — primarily bank, brokerage, and credit-card statements — and structures that data for review and analysis. It does not render professional opinions and is not a substitute for the judgment of a qualified forensic accountant or investigator.

The professional using the platform remains responsible for reviewing, verifying, and interpreting the extracted data and for any conclusions drawn from it. This document describes process; it is not legal advice. The admissibility and evidentiary weight of any data or analysis are determined by the court, not by eFraud Services Inc.

2. Governing Principle: Transcription, Not Computation

The single principle that governs the platform’s design is that it transcribes what appears on the source document and does not compute, estimate, infer, or “correct” the financial values it reports. Debits, credits, balances, dates, and check numbers are captured as printed.

This restraint is deliberate. A malformed, impossible, or internally inconsistent value may itself be evidence — of alteration, or of an irregularity in the underlying records. Silently normalizing such a value would destroy its evidentiary significance. The platform therefore preserves what is on the page and surfaces anomalies rather than smoothing them away.

General-purpose generative AI	eFraud Investigator's constrained use
Creates or infers content not present in the source	Locates and maps only content that appears on the page
May estimate, "fill in," or normalize values	Instructed not to estimate, fill, or normalize values
Interprets meaning and generates narrative	Transcribes only; interpretation is left to the professional
Acts as an author	Acts as a digital scribe under fixed system instructions

Through fixed system instructions, the model is directed to refrain from:

- Filling in missing or illegible figures, or estimating absent values.
- Adjusting figures based on surrounding context.
- "Smoothing" or correcting values it might judge to be incorrect.
- Producing any value that does not appear in the source document.

> ⚠ **An honest statement of what these instructions are**
>
> These are constraints on the model's instructions, not a hardware-level guarantee. Because the underlying model is generative, its output is not certified to be byte-for-byte identical on every run. The platform does not rely on the model behaving perfectly. It relies on the layered verification described in Sections 4 through 7 — source tie-out, reconciliation flagging, the tamper-evident audit trail, and the professional's own review — to surface and correct any error before the data is relied upon.

3. Transcribed Data vs. Derived Annotations

The platform distinguishes two kinds of fields, and the distinction matters for how each should be treated as evidence.

- Transcribed data — the amounts, dates, and identifiers captured from the source document. These are the evidentiary record and are verifiable against the source page.
- Derived annotations — analytical aids such as an inferred payee or counterparty. These are conveniences for the analyst, not transcribed source facts, and should be confirmed by the professional before being relied upon.

The platform does not automatically merge payee variants or collapse derived labels, precisely because doing so would substitute the system's judgment for the professional's on a point an opposing party could contest.

4. Source Verification: Per-Transaction Tie-Out

Each extracted transaction is associated with the specific source page it was drawn from. Within the application, the professional can view any extracted value against the original document image and confirm it against the source. No figure in the output is a black box: its provenance is available for inspection one step from the record.

This per-transaction tie-out is the primary mechanism by which an individual value is verified. It allows the professional — or an opposing expert — to test any data point directly against the document it came from.

5. Running-Balance Reconciliation: Corroboration, Not Proof

For each account, the platform computes a running balance across the extracted transactions and compares the result to the ending balance printed on the statement. When the two agree, this is meaningful corroboration that the transcribed debits, credits, and balances are internally consistent with the statement's own arithmetic. When they do not agree, the platform surfaces the discrepancy so the professional can locate and resolve it.

> ☑ **What a passing reconciliation shows**
>
> That the transcribed monetary values, taken together, reproduce the statement's own arithmetic. This is strong corroboration of the numeric transcription and the same internal-consistency check a professional would apply to a manually prepared schedule.

> ⚠ **What reconciliation does not establish**
>
> Reconciliation tests arithmetic, not every field. Two offsetting transcription errors can still reconcile, and non-arithmetic fields — dates, check numbers, payee names, and descriptions — are not tested by the balance at all. Reconciliation is therefore a corroborating check and a discrepancy flag, not a guarantee that every value is correct. Final verification of any individual value rests with the professional, using the source-document review described in Section 4.

A dataset that fails to reconcile is flagged for review. The professional may adopt the practice of requiring a dataset to reconcile before relying on it; the platform supports this as a review practice consistent with the expert-oversight model described in Section 7.

6. The Correction Audit Trail

Every correction a user makes to an extracted value is recorded in an append-only audit log capturing the original value, the substituted value, the user who made the change, and the time it was made. The log is hash-chained: each entry incorporates a cryptographic digest of the prior entry, so any later alteration or deletion of a recorded correction is detectable.

Provenance fields recording who uploaded and who created each record are written once, from the authenticated user's identity, and are not editable through the application. The result is a

tamper-evident chain of custody for every change made to the data after extraction — a record the testifying professional can stand behind when asked how the data reached its final state.

7. Human Oversight and Expert Sponsorship

The platform is built on the expert-oversight model that courts expect of instrument-assisted analysis: the tool performs the mechanical work; the qualified professional reviews, verifies, and sponsors the result. The platform provides the review capabilities below. Their use is the professional's responsibility and is not enforced by the software.

Capability	What it provides
Line-by-line review	Extracted data can be reviewed against the source before output is finalized or exported.
Manual correction	Where a value is misread, the professional corrects it directly; the platform does not make "best-guess" substitutions on illegible text.
Source tie-out	Any value can be audited against the original document image within the application (Section 4).
Audit trail	Every correction is logged in the tamper-evident record described in Section 6.
Expert certification	The testifying professional remains the sponsor of the evidence and certifies the dataset before use.

> **ℹ Where responsibility rests**
>
> These are tools the professional uses; the platform does not force review or block export. The evidentiary weight of the result rests on the professional's diligence — exactly as it would for a manually prepared schedule. This mirrors the standard applied to forensic laboratory instruments: the technology performs the measurement; the credentialed professional interprets and sponsors the result.

Compared with manual entry, machine transcription does not suffer the fatigue-related accuracy degradation that affects a person transcribing large volumes of figures. Combined with source tie-out and reconciliation flagging, this supports consistent handling of large datasets. It does not remove the need for professional review; it changes where the professional's time is spent — from typing to verification.

8. Reproducibility and Testing

The extraction pipeline is exercised against a regression test suite — a fixed set of representative source documents run end to end through the live pipeline — so that changes to the system can be checked for consistent results before release.

Because the underlying model is generative, the platform's reliability does not rest on byte-identical model output. It rests on the layered verification described in this memorandum: constrained transcription instructions, per-transaction source tie-out, reconciliation flagging, the tamper-evident audit trail, and professional review.

9. Relationship to the Rules of Evidence

The output is designed to function as a summary of voluminous records under Federal Rule of Evidence 1006 and its state analogues. Whether it is admitted, and what weight it carries, is determined by the court — not by eFraud Services Inc.

FRE 1006 consideration	How eFraud Investigator supports it
Originals available for examination	All output derives from the original source PDFs, which remain available for inspection by any party.
Summary fairly reflects the originals	Each value can be verified against its source page; reconciliation corroborates the internal arithmetic of the numeric data.
Transparent methodology	The process is a digital analog of manual data entry, with a tamper-evident record of every correction.
Qualified sponsoring witness	The professional reviews and certifies the dataset, maintaining the same foundation as a manually prepared schedule.

In weighing reliability, courts and qualified reviewers commonly consider whether a method can be and has been tested, whether it produces consistent results, whether errors can be detected and addressed, and whether the process follows defined standards. The methodology described here was built with those considerations in mind: testability through the regression suite (Section 8); error detection through source verification and reconciliation flagging (Sections 4 and 5); and defined standards through the fixed transcription instructions (Section 2) and the audit trail (Section 6). Whether a given method satisfies the legal standard for admissibility in a particular matter is a question for counsel and the court, and nothing here should be read as a representation on that question.

10. Scope and Limitations

- The platform is a transcription and organization tool. It does not render professional opinions and is not a substitute for a qualified professional's judgment.
- Derived annotations, such as payee, are analytical aids and should be confirmed before being relied upon.
- A passing reconciliation corroborates arithmetic consistency; it does not certify that every field is correct.
- Review and verification tools are provided to the professional; their use is the professional's responsibility and is not enforced by the software.

- This document describes process and is not legal advice. Admissibility and weight are for the court.

11. Summary

eFraud Investigator transcribes financial data from source documents without computing, estimating, or altering the values it reports. It links every transaction to its source for independent verification, corroborates the numeric data through running-balance reconciliation while flagging discrepancies for review, records every post-extraction correction in a tamper-evident audit trail, and is exercised against a regression suite for consistency.

It distinguishes transcribed source data from derived analytical annotations, and it leaves interpretation, verification, and professional judgment where they belong — with the qualified professional who sponsors the result.

eFraud Services Inc. • Naples, FL • efraudservices.com • June 2026