



**AutoForge AI**  
AUTOMATION WITH AI PRECISION

## **AutoForge Vision**

### **Modernizing Chain of Custody Processing for Environmental Laboratories**

**Company:** AutoForge AI LLC

**Product:** AutoForge Vision

**Status:** Development – Pilot Program Available

**Author:** James Phifer, Founder & CEO

## Executive Summary

Environmental laboratories continue to rely on handwritten Chain of Custody (COC) forms as the legal and operational foundation of sample intake. Manual transcription of these forms into Laboratory Information Management Systems (LIMS) remains a major bottleneck, introducing

- High labor costs
- Transcription errors
- Standard Operating Procedure (SOP) non-compliance risk
- Delays in sample processing

**AutoForge Vision** is a self-contained, air-gapped AI platform that converts handwritten COC forms into validated, LIMS-ready structured data using local Vision Language Models (VLMs), computer vision, and deterministic validation logic.

The system reduces manual labor per 50-sample COC from approximately **35 minutes to 5 minutes**, while enforcing structured validation prior to LIMS insertion.

## The Operational Problem

### Manual Intake Baseline

For a typical 50-sample COC:

- 20–30 minutes transcription
- 5–10 minutes QA review
- Approx. 35 minutes total labor per form

Industry-standard transcription error rates range from 3–5% for manual alphanumeric entry. Errors in sample IDs, matrices, hold times, or analysis requests can result in:

- Re-sampling
- Regulatory exposure
- Litigation risk
- Internal data repair cycles

Manual intake remains one of the most fragile points in the environmental lab workflow.

## AutoForge Vision Architecture Overview

AutoForge Vision is designed for secure, local, high-throughput operation.

### Deployment Model

- Docker containers
- Runs inside local Linux virtual machine
- No internet required for AI inference or data processing
- Fully air-gapped capable

The form is scanned at the lab, and the scanned file is retrieved from the local network or lab computer and uploaded into the AutoForge Vision app. All data remains securely within the lab's controlled environment.

### AI Processing Stack

AutoForge Vision uses a multi-model architecture optimized for structured form extraction.

### Vision and Extraction Layer

- **Vision Language Model**  
Primary vision-language model for contextual understanding of the entire form.
- **Object Detection Model #1**  
Table detection and structural segmentation.
- **Object Detection Model #2**  
Checkbox detection and matrix mapping.
- **OCR Model**  
Lab identification and header detection

This layered approach allows the system to interpret both structured and semistructured handwritten forms.

### Handwriting Accuracy

- Approx. 90%+ for normal handwriting.
- Approx. 80% for messy field handwriting.
- Accuracy dependent on image clarity and penmanship.
- Human verification closes the final gap.

## End-to-End Data Pipeline

The system follows a deterministic multi-stage workflow:

### 1. Intake

- Client uploads COC PDF into the Autoforge app
- COC form auto-detected via keyword matching
- Known lab rules applied when available
- Generic fallback rules used otherwise

### 2. Vision Processing

- PDF pages converted to high-resolution images
- Vision model extracts:
  - Project
  - COC Number
  - Address
  - Contact
- Object Detection Model identifies table rows and checkboxes
- Sample rows mapped to analysis requests

### 3. Data Construction

This step finalizes the structured record for each sample. Header data (including **Project**, **COC Number**, **Address**, and **Contact**), which was extracted in the Vision Processing stage, is merged with the individual sample rows. This process leverages the programmatic matching done previously to intelligently link the **Tests** (analysis codes) to the correct **Sample rows**. The result is a complete, per-sample structured record containing the following fields:

- Sample ID
- Matrix
- Date
- Containers
- Tests

### 4. Validation Gate

All records pass through deterministic validation, compliant with the **SOP**, before export. New validation rules can be programmatically added to the system as required by the **SOP** to ensure comprehensive compliance.

## Deterministic Validation Engine

AutoForge Vision enforces structured SOP logic prior to LIMS upload, ensuring regulatory compliance (Supports 21 CFR Part 11 compliance).

## Required Field Validation

Flags error if missing:

- Sample ID
- Matrix
- Comp/GRAB
- Number of containers
- Lab name
- Lab address
- Required analysis identifiers

## Composite Start and End

- Start time required
- End time required
- End time must be after start time
- System flags chronology errors.

## Uniqueness Enforcement

- All sample IDs must be unique within COC
- Prevents LIMS overwrite conflict

## Structural Validation

- Matrix codes validated
- Composite rules enforced
- Analysis mapping verified
- If errors found:
  - Errors are attached to specific sample
  - The errors are displayed in user interface
  - Correction is mandatory before final export

No data reaches LIMS without passing validation.

## Human-in-the-Loop Interface

Frontend built in **NextJS**.

Split-screen design:

- Right: Original COC image
- Left: Editable spreadsheet-style extracted data

Features:

- Real-time processing updates via Redis SSE
- As each form finishes extraction, it becomes available immediately
- Lab personnel can correct flagged errors
- All edits are logged

## Audit Trail and Compliance

The AutoForge Vision system maintains a complete digital audit log, recording:

- **User ID:** Authenticated via Google Sign-In (requires network access)
- **Timestamp**
- **Original extracted value**
- **Corrected value**

Compliance and documentation support includes:

- Supports 21 CFR Part 11 alignment and litigation defensibility.
- The original PDF image is attached to the LIMS Analysis Request.

## LIMS Integration

Supports:

- REST API push
- CSV export

On “Save” click:

- JSON transformed to LIMS-specific format
- Client and contact resolved or created
- Sample types and services mapped
- Analysis Request created
- Original COC attached

## GPU Infrastructure and Throughput Modeling

Minimum supported GPU:

- NVIDIA A100 (40GB VRAM)

### Measured A100 Performance

- GPU VRAM: 40 GB
- Approx. 100 pages/hour per GPU
- Batch size: 4 pages

### Measured H100 Performance

- GPU VRAM: 80 GB
- Expected throughput: 250–400 pages/hour per GPU
- Batch size: 6 pages

### Measured B200 Performance

- GPU VRAM: 192 GB
- Expected throughput: 1000–1500 pages/hour per GPU
- Batch size: 12 pages

#### Notes:

- The A100 GPU performance is based on observed testing; H100 and B200 throughput figures are estimates.
- Throughput scales with the addition of more GPUs, enabling faster parallel processing of forms.

Scaling depends on:

1. Compute
2. Memory bandwidth
3. GPU Utilization
4. CPU I/O Constraints
5. Model Architecture

## AI with Human in the Loop

- GPU and human verification operate concurrently.
- As soon as one form completes extraction, it becomes available for review while others continue processing.

This ensures:

1. No GPU idle time
2. No staff waiting
3. Balanced compute-to-human workflow

## Security and Data Residency

- Fully self-contained.
- No external API calls required for AI inference or data processing.
- The core pipeline operates fully air-gapped.
- Network access is required only for user authentication via Google Sign-In.
- **Optional fully air-gapped user authentication:** For environments requiring absolute isolation, Google Sign-In can be replaced with a local, self-contained username and password authentication system, ensuring the entire application stack is air-gapped.
- No cloud inference.
- No data leaves the local environment.
- Local Supabase database.
- Dockerized deployment.

The system can operate fully air-gapped.

## Strategic ROI

- **Manual processing:** 35 minutes per 50-sample COC
- **AutoForge Vision:** 2.5 minutes extraction + 2.5 minutes verification = 5 minutes total
- **Labor reduction:** 85% per form

Additional benefits:

1. Reduced transcription errors
2. Reduced rework cycles
3. Increased intake throughput
4. Stronger compliance posture
5. Improved employee satisfaction

## Current Status

AutoForge Vision is in active development.

We are seeking environmental laboratories to participate in a structured pilot program.

Pilot labs will:

- Influence validation rule configuration
- Integrate into their LIMS environment
- Evaluate performance under real intake conditions

## About AutoForge AI LLC

AutoForge AI LLC develops secure, high-performance AI systems for industrial and scientific environments.

### **Our mission:**

Deliver enterprise-grade AI capabilities without exposing critical data to public cloud infrastructure.

## Contact

### **James Phifer**

Founder & CEO

AutoForge AI LLC

Email: [james.phifer@autoforgeai.ai](mailto:james.phifer@autoforgeai.ai)

Website: <https://autoforgeai.ai>

Phone: 864.320.2265