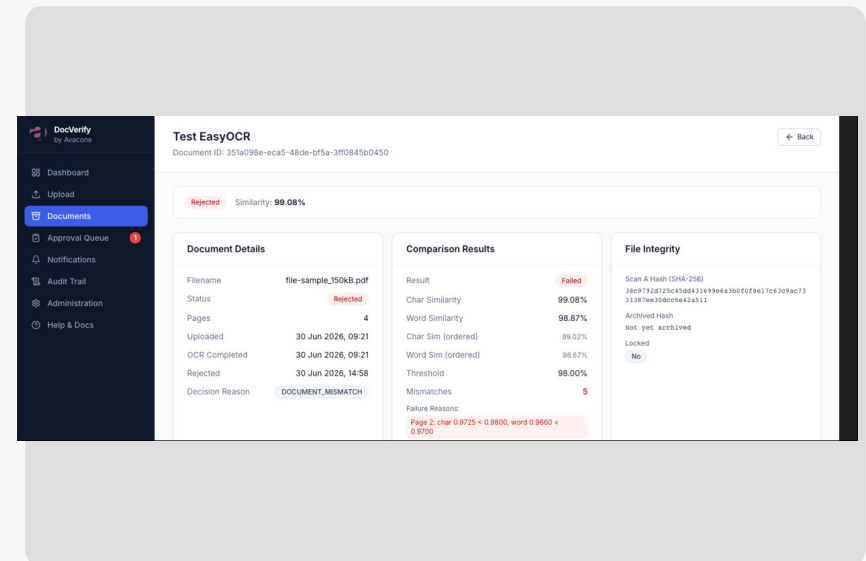


DocVerify

Audit-ready document verification for regulated pharmaceutical teams.

Controlled OCR. Deterministic comparison. Immutable evidence.



OCR alone does not create audit-ready evidence.

Pharma teams are not only checking whether text was extracted. They must prove what was compared, what changed, who reviewed it, and why the decision was defensible.

01

Manual verification does not scale

Hundreds of pages, binders, scans and supplier documents create review load that is too slow and inconsistent for manual checking.

02

Small mismatches carry risk

Batch records, QA forms and SOP evidence can fail on small differences that a simple pass/fail label may hide.

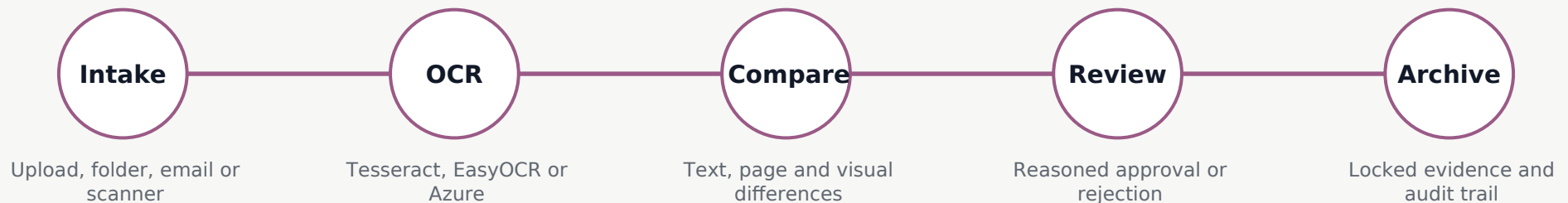
03

Evidence must survive inspection

Regulated buyers need reviewer decisions, source artifacts, file hashes and a traceable audit chain.

A governed verification layer, not another upload tool.

DocVerify turns document comparison into a controlled system of record with configurable OCR, reviewer evidence, immutable storage, and enterprise identity.



Buyer promise: fewer ad hoc checks, stronger evidence, and a clearer inspection posture.

The system adapts to the document source.

One platform supports single-file confidence validation, two-scan comparison, and scanner-driven intake workflows.

Mode A

Single-file dual OCR

One source file is processed twice with different OCR configurations to surface extraction instability.

Mode B

Two-file comparison

Two files or two scanner captures are processed with the same OCR strategy and compared side by side.

Mode C

Scanner-driven intake

Manual uploads, monitored folders, IMAP ingestion and scanner sources feed one controlled review pipeline.

From scan intake to audit-ready decision.

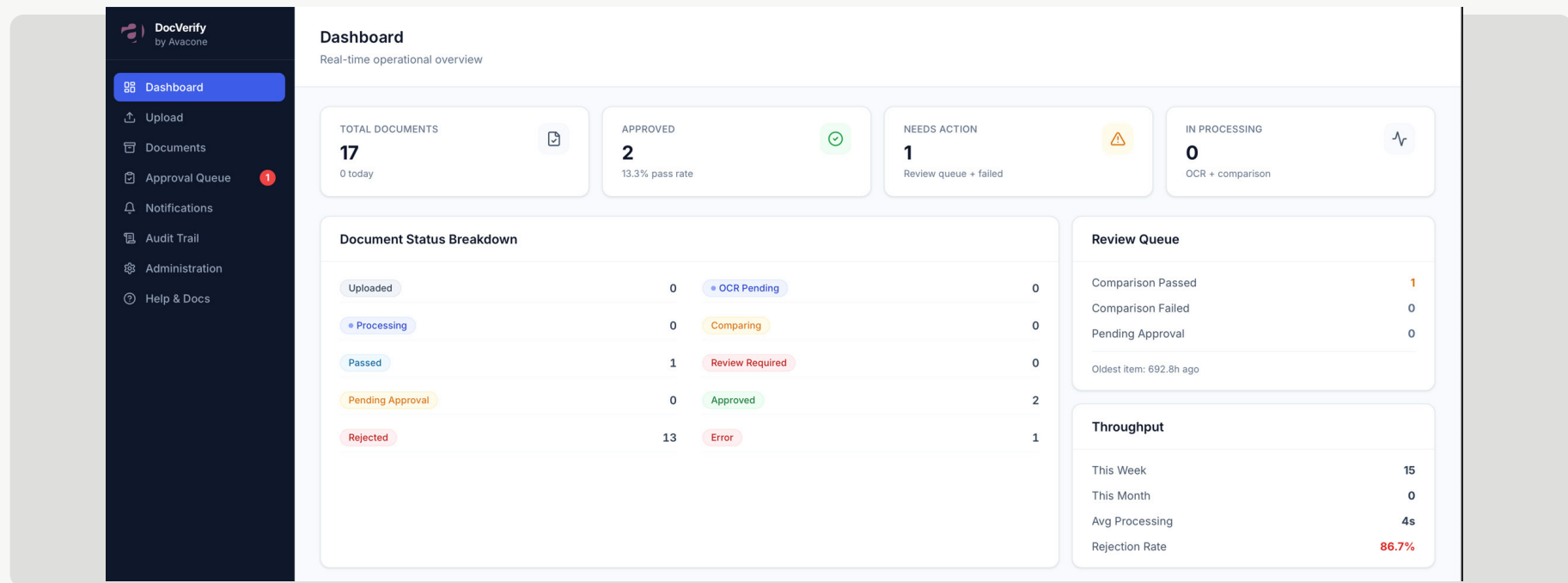
Automated processing does the heavy lifting while uncertainty is routed to human judgement.



The workflow is conservative by design: uncertainty is made visible instead of hidden inside a score.

An operational center for QA, scanning, review, and approval teams.

Fast status scanning, controlled queues and direct access to the evidence needed for defensible decisions.



Real-time dashboard

Structured intake

Document register

Review queue

No black-box score. Every difference is inspectable.

Reviewers can move from machine scoring to side-by-side evidence before making the final decision.

Original

Human visual verification

OCR text

Full extracted content

Differences

Changed, removed or added text

Visual diff

Page-image highlighting

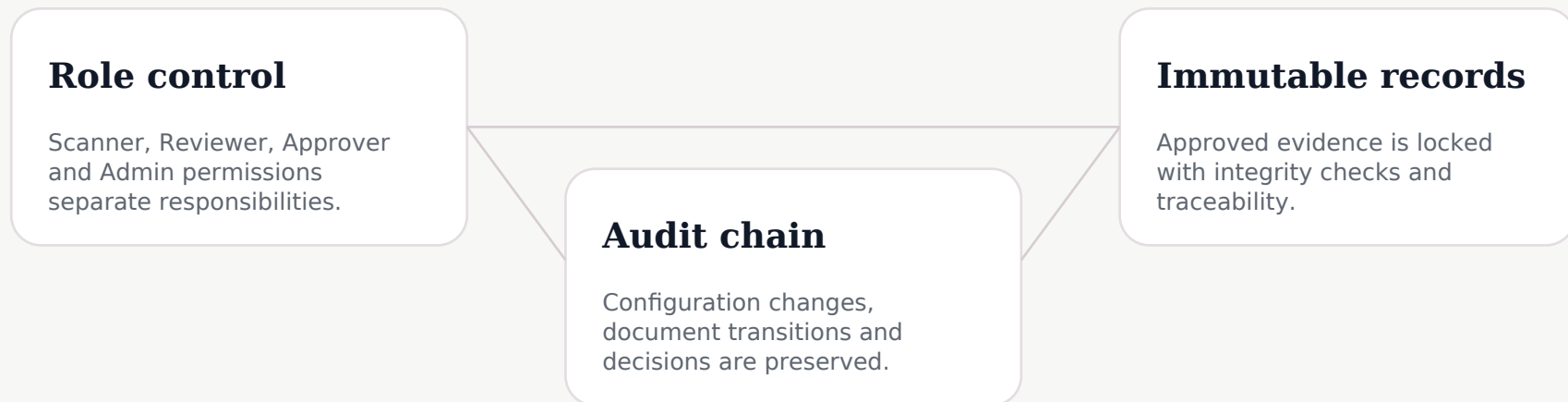
The screenshot displays the DocVerify by Avacone interface. On the left is a dark sidebar with navigation links: Dashboard, Upload, Documents (active), Approval Queue (with a red notification badge), Notifications, Audit Trail, Administration, and Help & Docs. The main area shows a comparison of two document scans, labeled 'SCAN A — PAGE 2' and 'SCAN B — PAGE 2'. At the top, it indicates 'Page 2 of 4' and '1 / 22 region(s) - engine: text layer'. A message states: 'Highlighted regions differ from the other scan (each side shows its own differences)'. Both scans show a table with 5 rows and 3 columns. The table content is as follows:

	Lorem ipsum	Lorem ipsum	Lorem ipsum
1	In eleifend velit vitae libero sollicitudin euismod.	Lorem	Lorem
2	Cras fringilla ipsum magna, in fringilla dui commodo ipsum		
3	Aliquam erat volutpat.	Lorem	
4	Fusce vitae vestibulum velit.	Lorem	
5	Etiam vehicula luctus fermentum.	Ipsum	

Below the table, there is a paragraph of text. In the comparison view, certain words and phrases are highlighted in red in both scans, indicating differences between the two versions of the document.

Built around accountability and audit readiness.

DocVerify is a controlled workflow layer for regulated decisions: identity, roles, configuration, locked records and evidence that can be inspected later.



Evidence is searchable, configurable, and verifiable.

Audit and administration views give quality and IT teams the controls needed for regulated operation.

Audit Trail
Complete append-only record of all system actions

Document ID: Filter

Event Type: Username: Date From: Date To:

#	EVENT	USER	DOCUMENT	DETAILS	TIMESTAMP	HASH
394	user.login	admin	—	—	01 Jul 2026, 10:37	596babab8...
393	user.login	admin	—	—	01 Jul 2026, 10:32	28a9d74d...
392	user.login	admin	—	—	01 Jul 2026, 10:32	f444a379...
391	user.login	admin	—	—	01 Jul 2026, 10:30	8f29a455...
390	config.changed	admin	—	ldap — enabled: on → off	01 Jul 2026, 10:26	8896c9a3...
389	config.changed	admin	—	ldap — url: sad → a	01 Jul 2026, 10:26	ae4a0616...
388	config.changed	admin	—	ldap	01 Jul 2026, 10:25	f4247611...
387	config.changed	admin	—	ldap — url: a → sad	01 Jul 2026, 10:25	d904d2b0...
386	config.changed	admin	—	ldap — enabled: off → on	01 Jul 2026, 10:25	ba520979...
385	config.changed	admin	—	ldap	01 Jul 2026, 10:25	8b80b5f8...
384	system.startup	system	—	—	01 Jul 2026, 10:15	085128db...

Audit trail

Search event history and verify the integrity of the audit chain.

Administration
User management, system configuration, and license status

Users Scanner Sources Categories **Settings** Console License

OCR & Engine
Engine, language, render quality, confidence

OCR Processing
Controls OCR language, image render quality (DPI), and whether OCR passes A and B run simultaneously. Higher DPI improves accuracy but increases processing time.

OCR quality note
OCR output depends on the engine, scan quality, selected language, and render DPI. German + English with moderate DPI suits most Novartis-style forms. The settings below apply to whichever engine is active - the same render DPI and language carry over.

OCR engine

☐ Tesseract Default
Fastest, lowest resource use. Best for clear, high-quality printed German/English scans and routine high-volume processing.

☒ EasyOCR Optional
Neural detection + recognition. Choose for degraded, low-contrast, or unusual-font scans where Tesseract struggles. More robust on hard inputs, but slower. Runs on both scans, so TrueCopy verification works identically.

☐ Azure Document Intelligence Optional - Cloud/Container **Needs credentials**
Microsoft's OCR service — strongest on handwriting and complex formats/templates. Requires the client's own Azure resource (cloud) or a local Azure AI container. Billing, if any, is between the client and Microsoft. Enter the Azure endpoint and API key in the Azure Document Intelligence panel below, then it becomes selectable here.

Azure Document Intelligence credentials **Needs credentials**

Comparison
Strategy, thresholds, blank & image pages

Automation
Automatic approval rules

System Admin

Administration

Configure OCR, comparison rules, scanner sources, users and reason codes.

Designed for years of binders and controlled local deployment.

DocVerify supports high-volume archives, sharded storage, background processing and on-premise operation inside the customer environment.

Binder-scale archives

Sharded paths keep large document populations operationally manageable.

Large files

Multi-hundred-page binders are handled through background jobs and page-level scoring.

Configurable throughput

Processing can run sequentially or in parallel depending on hardware and validation needs.

Enterprise identity

AD/LDAP aligns access with corporate identity and user lifecycle policies.

A product built for regulated buyers, not generic scanning.

The value is not OCR alone. The value is controlled review, defensible evidence, local deployment and long-term operational traceability.

Lower manual verification burden

Teams move from ad hoc visual checking to prioritized exception review.

Stronger inspection posture

Auditors can trace what entered the system, how it was processed, who reviewed it and whether the record remains intact.

Enterprise operating model

Deploy per site or centrally, tune to hardware, integrate identity and support high-volume retention.

Implementation stack for IT and validation teams.

The following technical components support the controlled OCR, comparison, workflow and audit model.

Frontend

React, TypeScript, Vite, Tailwind CSS

Backend API

FastAPI, Python 3.12, JWT, typed schemas

OCR engines

Tesseract, EasyOCR, Azure AI Document Intelligence

Processing

Celery workers, Redis queueing, sequential or parallel execution

Data layer

PostgreSQL 16, SQLAlchemy, Alembic, indexed search

Storage

Shard-based files, SHA-256 hashes, immutable approved records

Security and audit

RBAC, AD/LDAP, append-only audit events, hash-chain verification

Comparison

Character, word, page and visual difference evaluation

DocVerify

Controlled document verification for regulated pharma operations.

Recommended next step

Run a controlled pilot with representative batch records, QA forms, SOP evidence and scanned binder material to validate OCR behavior, comparison quality and review workflow fit.

On-premise deployment

Inspectable evidence

Controlled review workflow

Immutable audit-ready records