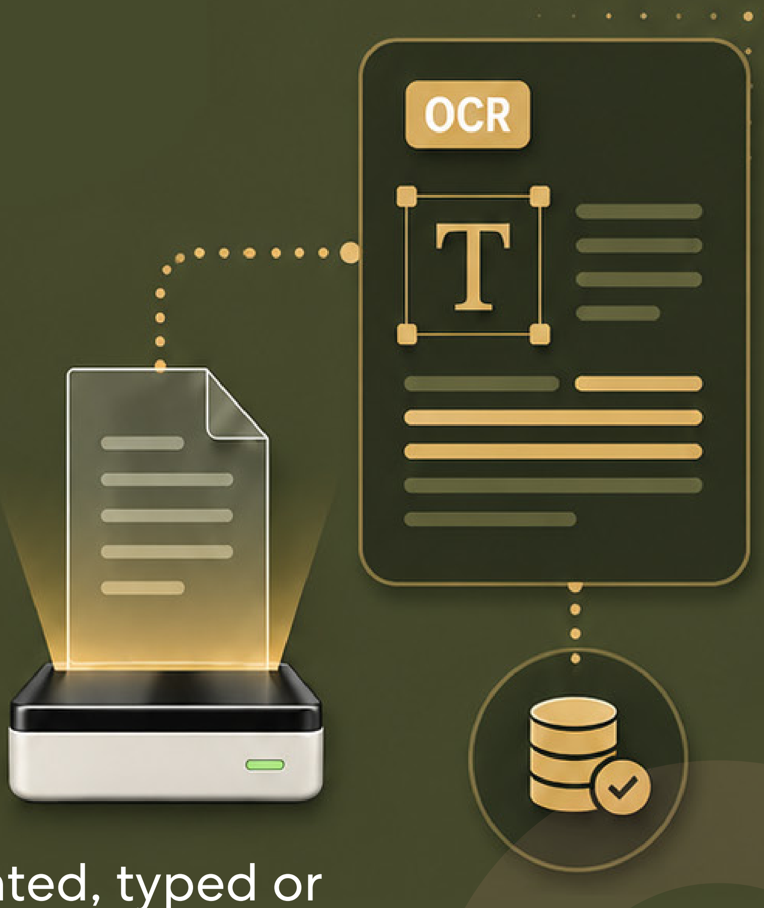
Shaip
by ubiquity

OPTICAL CHARACTER RECOGNITION

OCR in 2026



Turning any document - printed, typed or handwritten - into structured, machine-readable data that AI can understand and act on.

$55.3B	Projected OCR market by 2033
15.5%	Annual growth (CAGR)
98–99%	Page-level accuracy

What is OCR?

The field, in one line

Optical Character Recognition (OCR) converts images of text scans, photos, PDFs, handwriting into machine-readable, editable, searchable data. Modern OCR goes further: it doesn't just read the characters, it extracts, structures, and validates the information.

OCR vs IDP

OCR reads the text. **Intelligent Document Processing (IDP)** understands the document and acts on it - classifying, extracting fields, validating & routing.

80% of business data lives in **unstructured documents** - the OCR paperwork unlocks

How OCR Works

The modern 6-step pipeline

- **Capture** Scan/photograph
- **Pre-process** De-skew & clean
- **Detect Text** Locate regions
- **Recognize** AI reads chars
- **Structure** Fields & tables
- **Validate** Confidence+review

Types of OCR

Not all OCR is the same

OCR	Printed & typed text
ICR	Handwritten characters
IWR	Handwritten Words
OMR	Checkboxes & marks
IDP / AI-OCR	Understands & acts

OCR in the AI Era

Raw OCR is now just step one

Intelligent Document Processing
Classifies, extracts, validates & routes data end-to-end.

AI / Deep-Learning OCR
Generalizes across layouts — no brittle templates.

ICR — Handwriting
Reads cursive and handwritten notes.

NLP Understanding
Interprets meaning, not just characters.

LLM & Multimodal AI
Reads a page like a human: text, tables & images.

Agentic Workflows
Extract → validate → file, on their own.

Confidence Scoring & Human-in-the-Loop
Automation with targeted human review where confidence is low or risk is high.

OCR Use Cases

Where it delivers value

OCR INDUSTRY

Banking & Finance: Invoice & AP automation · KYC / identity onboarding · bank-statement & cheque processing · loan documents

Insurance Multi-document claims processing with exception handling · policy document extraction

Healthcare Digitizing handwritten notes & records · lab reports · prescription & medical-label reading

Logistics Bills of lading · proof of delivery · customs & shipment documentation

Government Passport / ID verification · forms digitization · records archives

Retail & Operations Receipt & expense processing · contract & vendor document extraction

Benefits & Challenges

The honest picture

Benefits

- ✓ Eliminates manual data entry
- ✓ Faster processing & turnaround
- ✓ Searchable, editable, accessible documents
- ✓ High accuracy 98-99% at page level
- ✓ Scales across document types & teams

Challenges

- ⚠ Poor-quality or noisy scans
- ⚠ Handwriting & unusual fonts
- ⚠ Changing / complex layouts (tables, stamps)
- ⚠ Validation, governance & audit trails
- ⚠ Integration with ERP / CRM systems

The OCR Market

Sourced & current

$55.3B
Growing 15.5% CAGR
Projected global OCR market by **2033**, up from **$13.1B** in 2023.

98–99%
page-level accuracy (up to 99.5% on financial docs)

~21%
BFSI is the largest OCR vertical

Source: Market.us — OCR Technology Market, 2026 (verify latest figure before publishing).

What's Next: OCR Trends for 2026

The road ahead

Agentic IDP	End-to-end, minimal supervision
LLM Extraction	Any layout, no templates
Multimodal AI	Text, tables & images together
Handwriting / ICR	Higher cursive accuracy
Straight-Through	Touchless automation

www.shaip.com

Sources: Market.us | Devox Software | MarketsandMarkets

Shaip
by ubiquity